Key Points:
- Observations are utilized to constrain cloud-system resolving model to improve the prediction of continental shallow cumulus populations
- The impact of assimilating boundary layer measurements on simulation of shallow cloud populations is especially examined
- The evolution of shallow clouds is better simulated as more realistic ambient environment and boundary layer are represented in the model

Correspondence to:
S.-L. Tai,
sheng-lun.tai@pnnl.gov

Simulation of Continental Shallow Cumulus Populations Using an Observation-Constrained Cloud-System Resolving Model

Sheng-Lun Tai[1] [ID], Jerome D. Fast[1] [ID], William I. Gustafson Jr[1] [ID], Duli Chand[1] [ID], Brian Gaudet[1] [ID], Zhe Feng[1] [ID], and Rob Newsom[1] [ID]

[1]Pacific Northwest National Laboratory, Richland, WA, USA

Abstract Continental shallow cumulus (ShCu) clouds observed on 30 August 2016 during the Holistic Interactions of Shallow Clouds, Aerosols, and Land-Ecosystems (HI-SCALE) field campaign are simulated by using an observation-constrained cloud-system resolving model. On this day, ShCu forms over Oklahoma and southern Kansas and some of these clouds transition to deeper, precipitating convection during the afternoon. We apply a four-dimensional ensemble-variational (4DEnVar) hybrid technique in the Community Gridpoint Statistical Interpolation (GSI) system to assimilate operational data sets and unique boundary layer measurements including a Raman lidar, radar wind profilers, radiosondes, and surface stations collected by the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) atmospheric observatory into the Weather Research and Forecasting (WRF) model to ascertain how improved environmental conditions can influence forecasts of ShCu populations and the transition to deeper convection. Independent observations from aircraft, satellite, as well as ARM's remote sensors are used to evaluate model performance in different aspects. Several model experiments are conducted to identify the impact of data assimilation (DA) on the prediction of clouds evolution. The analyses indicate that ShCu populations are more accurately reproduced after DA in terms of cloud initiation time and cloud base height, which can be attributed to an improved representation of the ambient meteorological conditions and the convective boundary layer. Extending the assimilation to 18 UTC (local noon) also improved the simulation of shallow-to-deep transitions of convective clouds.

Plain Language Summary Accurate prediction of life cycle of shallow convective clouds is very challenging for the existing weather and climate models since they have difficulties in reproducing realistic atmospheric structure within a shallow layer near the Earth surface (also called boundary layer, roughly below 2-km height in daytime). To tackle this fundamental problem, the observational data collected for operational weather prediction as well as unique boundary layer observations measured near north-central Oklahoma are integrated to constrain the behavior of cloud-system resolving model across different scales with an emphasis on boundary layer. The results show that the model biases in atmospheric conditions, especially humidity within boundary layer, are reduced with the modification informed by observations. As a consequence, shallow convective clouds are well reproduced in terms of cloud evolution in time and space which are verified by various cloud measurements. It also suggests that surface observation can be used to correct cold pool intensity which is closely related to the maintenance of deep convective clouds that are transitioned from shallow convective clouds.

1. Introduction

While individual shallow cumulus (ShCu) clouds are small, populations of ShCu that form over continents and trade wind regions over oceans alter atmospheric stability by redistributing heat and moisture and significantly affect the Earth's radiative budget (Berg et al., 2011; Neggers et al., 2007). ShCu forms when thermals gain enough buoyancy from within the boundary layer and overshoot the lifting condensation level (LCL). The diameter of ShCu is generally less than 1 km and has a depth usually smaller than 2 km (Berg & Kassianov, 2008). It is also called "fair-weather" cumulus since it rarely precipitates or produces very light rainfall. The life cycle of ShCu is very sensitive to local turbulence and updrafts within boundary layer convective eddies (Lareau et al., 2018). Under certain environmental conditions, ShCu can transition into deeper, precipitating convection and ultimately more organized deep convective systems

(Khairoutdinov & Randall, 2006; Kuang & Bretherton, 2006; Wu et al., 2009; Zhang & Klein, 2010, 2012). Thus, improving prediction of shallow convective cloud populations would benefit not only solar radiation estimates for climate modeling and solar energy forecasts (Jimenez et al., 2016), but also estimates of precipitation amount and frequency that are crucial for climate, weather forecasting, and water resources management.

To better understand and characterize the processes influencing the life cycle of continental ShCu and other atmospheric phenomena, extensive long-term measurements have been obtained at the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) atmospheric observatory. The SGP site, located in north-central Oklahoma, is well suited to study ShCu (Berg et al., 2013; Berg & Kassianov, 2008; Wagner et al., 2013; Zhang & Klein, 2010, 2012) because of the high climatological occurrence of locally forced ShCu in the region. Several field campaigns have also been conducted around the SGP site to investigate processes affecting the life cycle of shallow clouds, including the Holistic Interactions of Shallow Clouds, Aerosols, and Land-Ecosystems (HI-SCALE) field campaign (Fast et al., 2018) conducted during the spring and summer of 2016. The spring and summer Intensive Observation Periods (IOPs) occurred between 24 April to 21 May and 28 August to 23 September, respectively. Processes connecting the soil, surface layer, boundary layer, and free troposphere can be investigated by integrating the HI-SCALE measurements made by the Gulfstream I (G-1) aircraft (Schmid et al., 2013), the ARM ground instrumentation (e.g., scanning and vertically pointing radars, radiosondes, Doppler lidars (DLs), wind profilers, radiometers, surface meteorology, eddy correlation systems, soil temperature, and moisture) (Sisterson et al., 2016), and the Oklahoma Mesonet (e.g., surface meteorology, soil temperature, and moisture) (McPherson et al., 2007). These comprehensive data sets provide a wealth of information for detailed observational analysis, but also provide a valuable source of data to verify and improve high spatial resolution numerical simulations.

Uncertainties in the initial or boundary environmental conditions are a prominent source of error in numerical cloud forecasts. The research community frequently uses global reanalysis products produced by multiple agencies and countries to provide the initial conditions for global climate models and the initial and boundary conditions of regional cloud-system resolving models. These reanalyses are usually obtained from a global model constrained by conventional observations using data assimilation (DA) techniques. However, cloud-scale features are poorly resolved in these analyses because of the coarse grid spacing (both horizontal and vertical) in the host global model. The reanalyses may also contain larger uncertainties in regions with few valid observations to constrain the model. Numerous studies over the past two decades have also demonstrated the benefits of DA techniques that integrate the available observations with a cloud-system resolving models (Anderson et al., 2009; Barker et al., 2012; Benjamin et al., 2016; Hu et al., 2017; Meng & Zhang, 2008; Schwartz et al., 2015). Many studies have examined how DA influences the timing, location, spatial extent, and severity of deep convective systems (Chang et al., 2015; Johnson et al., 2015; Sun & Wang, 2012; Tai et al., 2011, 2017) since those events are hazardous and have immediate societal impacts.

In contrast, there have been far fewer studies focusing on the impact of DA on simulating more typical and benign weather conditions, including the life cycle of shallow convective clouds and their transition to deeper, precipitating convection. Since the formation and growth of shallow convective clouds are highly related to processes taking place within or near the convective boundary layer (CBL) (Lareau et al., 2018; LeMone et al., 2013), a realistic representation of CBL structures in the model is needed. Nevertheless, it has been found that significant model biases occur within the boundary layer. For instance, Morcrette et al. (2018) demonstrate extensive and significant temperature biases over the central United States in multiple weather and climate model forecasts. To tackle this well-known issue, numerous studies have used various DA techniques to assimilate meteorological observations at the surface and within the boundary layer (Adam et al., 2016; Alapaty et al., 2001; Ruggiero et al., 1996; Stauffer et al., 1990; Wulfmeyer et al., 2006). However, most of them do not assess the impact of assimilating these data on subsequent cloud prediction. Li et al. (2015) applied the Community Gridpoint Statistical Interpolation (GSI) system to assimilate the National Centers for Environmental Prediction (NCEP) operational data stream along with some of the ARM measurements by using a three-dimensional variational (3DVar) technique. Their multiscale DA (MSDA) technique essentially assimilates data with flexible scaling factors that can be tuned with respect to model resolved scales. The analysis generated by MSDA has been adopted by the DOE's Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) project that conducts
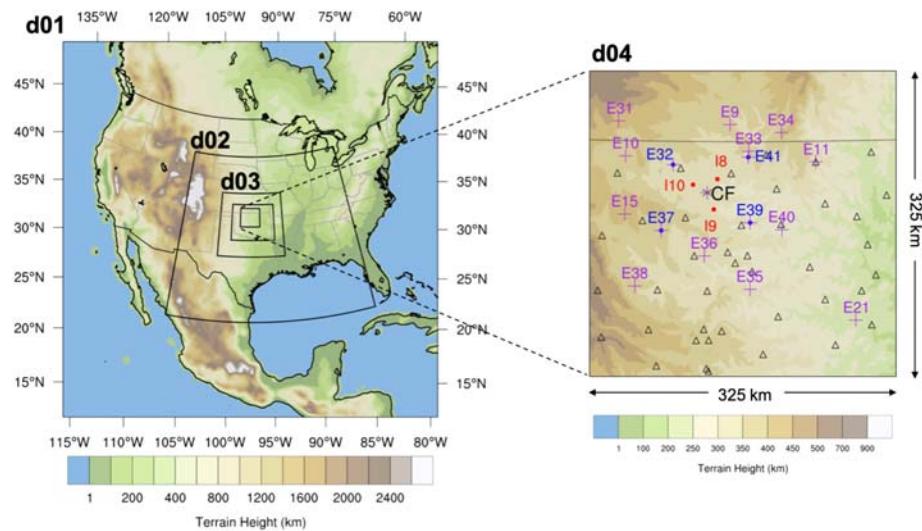
Figure 1. Geographic maps with four WRF model nesting domains (d01–d04) depicted. Color shading represents terrain height. Domain 4 is zoomed in to marks available observations including Central Facility, radar wind profiler sites (shown in red texts), Doppler lidars (blue), ARM surface sites (magenta), and Oklahoma Mesonet (empty triangles).

semi-operational high-resolution simulations of ShCu (Gustafson et al., 2020). Nevertheless, the MSDA's native prediction of ShCu populations has not been examined because it has only been used to provide the large-scale forcing of LES.

In contrast to these earlier studies, the goal of this research is to determine whether the predictability of ShCu cloud populations in a cloud-system resolving model can be improved when constrained by observations of environmental states collected in and around the ARM SGP site during a particular event in the HI-SCALE campaign. We apply a hybrid ensemble variational scheme to assimilate conventional as well as campaign observations to examine the impacts they have upon the simulated life cycle of shallow convective cloud populations including the transition to deeper convection. Overall, the results demonstrate a positive impact of DA on constraining multiscale meteorological conditions of the ambient environment, especially within the boundary layer, in the cloud-system resolving model, leading to more accurate prediction of shallow convective clouds evolution. An extended assimilation at 18 UTC shows positive impact on the simulation of shallow-to-deep transitions of convective clouds. The research tools employed, including the forecast model, DA techniques, and the observational data sets, are introduced in section 2. The experimental design and demonstration of a case study day containing a complete life cycle of ShCu clouds are described in section 3. In section 4, the simulated clouds are examined qualitatively and verified quantitatively across multiple spatial scales by comparison against satellite retrievals as well as ground-based remote sensors. Corresponding meteorological conditions are also evaluated by independent field campaign data to clarify what specific optimizations are done by DA. In section 5, we discuss some challenges and path toward improved modeling of shallow convective cloud populations, including challenges in representing transitions of shallow-to-deep convection, the necessity of hydrometeor assimilation, the frequency updates in operational forecast products, and the potential of mesoscale DA in LES modeling. Finally, a summary is included in section 6.

2. Research Tools and Data Sets

2.1. Forecast Model and DA Technique

The Weather Research and Forecasting (WRF) model version 3.9.1 (ARW, Skamarock et al., 2008) is used to conduct all the mesoscale cloud simulations in this study. Figure 1 depicts four one-way nested domains. The first domain, denoted as d01, has a grid spacing of 36 km and encompasses the continental United States, Canada, and adjacent oceans. Each inner domain's horizontal grid spacing decreases from its parent domain by a factor of 3 so that the grid spacings on the second, third, and fourth domains are 12, 4, and

1.33 km, respectively. The fourth domain (d04) encompasses most of the ARM SGP measurement sites as shown in the right panel of Figure 1. A stretched grid is used in the vertical direction with 74 levels. To better resolve the boundary layer and shallow clouds, smaller vertical grid spacings are used in the lower troposphere. For instance, the vertical grid spacing stretches from approximately 30 m near the surface to 60 m near 2-km height and then gradually increased to 300 m at altitude of 6 km. The simulations adopt the Morrison microphysics parameterization (Morrison et al., 2005), Mellor-Yamada-Janjic (MYJ) boundary layer parameterization (Janjić, 1994), MYJ surface layer parameterization (Janjić, 2001), Unified Noah land-surface parameterization (Chen & Dudhia, 2001), and the Rapid Radiative Transfer Model for General Circulation Models (RRTMG) longwave and shortwave radiation parameterization (Iacono et al., 2008). The cumulus potential (CuP) shallow convection coupled with Kain-Fritsch deep convection parameterization (Berg et al., 2013) is used only for the first and second domains.

All simulations are initialized with the NCEP FNL Operational Model Global Tropospheric Analysis (http://dss.ucar.edu/datasets/ds083.2/), which has atmospheric and soil variables on a $1 \times 1°$ grid. Analyses at 6-h intervals provide the lateral boundary conditions for the first WRF domain (d01). The land use data are obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS)-based data set available at a 1-km grid spacing with the International Geosphere-Biosphere Programme (IGBP) land cover type classification. These and other static water, land, and soil parameters are interpolated to the model domains using the WRF Preprocessing System (WPS).

The version 3.6 Community GSI system, utilized in this study, is capable of assimilating a wide range of observational data including conventional (e.g., radiosonde, wind profiler, land surface, buoy, radar velocity-azimuth display [VAD] algorithm wind profile, aircraft) and satellite radiance and retrieved properties (Shao et al., 2015). The system provides several widely used assimilation technique options, including 3DVar, three-dimensional and four-dimensional ensemble-variational hybrid (3DEnVar, 4DEnVar; Hamill et al., 2011; Wang et al., 2013; Wang & Lei, 2014), as well as the ensemble Kalman filter (EnKF; Zhu et al., 2013).

The 3DVar method uses static background errors generated climatologically from more than one month forecast data set by using NMC method (forecast minus analysis). Since these background errors are precalculated, 3DVar has the least computational cost among all DA techniques. However, from what was revealed in earlier studies (Bannister, 2017; Wang & Lei, 2014) and by our own preliminary test experiments, the ensemble-variational DA techniques (3DEnVar and 4DEnVar) overall outperform 3DVar in many occasions. Wang and Lei (2014) noted that the 3DEnVar technique does not account for the temporal evolution of the error covariance within the assimilation window, and only the ensemble perturbation at a single time period (the center of the assimilation window in this study) is used in calculation of the cost function during variational minimization, whereas the 4DEnVar technique can use information extracted from multiple ensemble perturbations at time periods within the assimilation window to generate time-evolving background error structures. EnVar methods are found to be more robust than EnKF when ensemble size is relatively small or when model errors are large as the variational method employs dynamic constraint during minimization. Based on these superior traits, we use the 4DEnVar hybrid technique in the current study. In comparison to 4DEnVar, 4DEnVar hybrid technique particularly blends ensemble-based background error with static background error to alleviate possible underestimation of errors represented by limited number of ensemble members. More details about the implementation and formulation of 4DEnVar can be found in Wang and Lei (2014).

To enable ensemble-variational hybrid DA, a group of ensemble members is essential as they comprise the "flow-dependent" background errors which accounts for 85% of the total background errors used in our study. The static background error is computed using forecasts from the NCEP's North American Mesoscale Forecast System (NAM) model covering North America and responsible for another 15% of total background error. The ensemble member forecasts run with the identical domain configuration (numbers, sizes, and resolutions) as that of the DA experiments have. Instead of perturbing the model on our own, the NCEP Global Ensemble Forecast System (GEFS) 21-member ensemble with 1° horizontal grid spacing (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-ensemble-forecast-system-gefs) is used to provide a set of candidates of initial and boundary conditions for high-resolution ensemble forecasts. To obtain reasonable ensemble spread, a size of 54 members is designed as each member is

Table 1
List of Parameterization Schemes Included in Generating Multiphysics Ensemble That are Used in Estimation of Flow-Dependent Background Error Covariance for Four-Dimensional Ensemble-Variational Hybrid Assimilation

Type of parameterization scheme	PBL	Cumulus (only for d01, d02)	Microphysics	Land surface
Choice no.1	MYJ	KF_CuP	Morrison	Noah
Choice no.2	YSU	Grell	Thompson	NoahMP
Choice no.3	MYNN	Tidtke	WSM6	None

initialized at 00 UTC on 29 August by one randomly selected GEFS member out of 21. Then, each simulation is integrated by 2 days and output every hour with an unrepeated combination of planetary boundary layer (PBL), cumulus, microphysics, and land surface parameterizations. Aside from those parameterization adopted in the default model configuration mentioned in section 2.1, Table 1 denotes two additional PBL parameterizations including the Yonsei University (YSU, Hong et al., 2006) and Mellor-Yamada-Nakanishi-Niino (MYNN, Hong et al., 2006; Nakanishi & Niino, 2009), two additional cumulus parameterizations including the Grell 3D ensemble (Grell & Dévényi, 2002) and Tiedtke (Tiedtke, 1989), two additional microphysics parameterizations including the Thompson (Thompson et al., 2008) and WRF single-moment 6-class (Hong & Lim, 2006), and one additional land surface parameterization (Noah-MP; Niu et al., 2011) used for the ensembles.

2.2. Observational Data Sets

Various observational data sets are used for DA and forecast evaluation. To ensure that the regional model is constrained by realistic large-scale atmospheric conditions, the NCEP ADP Global Upper Air and Surface Weather Observations data set (https://rda.ucar.edu/datasets/ds337.0/) and Global Data Assimilation System (GDAS) satellite data (https://rda.ucar.edu/datasets/ds735.0/) are assimilated. They consist of global upper air and surface weather observations as well as satellite data that are operationally used by the NCEP operational GDAS. The upper air and surface weather observations measured by multiple instruments (radiosonde, aircraft, ship, buoy, wind profiler, surface station, etc.) provide atmospheric states of pressure, temperature, wind, and humidity over the whole globe. However, the observations are distributed unevenly in space and concentrate in or near the continents except for ship and aircraft measurements. In contrast, the advantage of satellite data, such as brightness temperature and radiance, is its spatial coverage that provides information to fill in data voids at particular periods of time. Since the majority of these observations are located above the boundary layer, the synoptic environment is best improved by the NCEP GDAS data sets.

In addition to operational data sets, the surface and boundary layer atmospheric states measured by instruments deployed at the ARM SGP site and over Oklahoma are also incorporated to further constrain the atmospheric conditions in the model. The ARM program collects a wide range of measurements to support research that improves the basic understanding of the representation of clouds and radiative forcing in global climate models (Sisterson et al., 2016). There are currently 19 measurement sites over north-central Oklahoma, including the Central Facility that has the most extensive suite of instruments. The exact locations of the sites are depicted over the fourth model domain in Figure 1. ARM measurements used for assimilation in this study are radiosondes, radar wind profilers (RWPs), a Raman lidar (RL), and surface sites. The radiosondes are launched at the Central Facility every 6 h, providing vertical profiles of wind speed, wind direction, pressure, humidity, and temperature from the surface to the tropopause. Three RWPs deployed around the Central Facility measure the variability of lower tropospheric wind profiles around the SGP site every 10 min. The RL located at the Central Facility uses a number of narrow-band detection channels specifically tuned to sense the Raman backscatter from atmospheric N_2, O_2, and H_2O molecules. Moisture and temperature profiles are retrieved by combining raw signals from these channels (Newsom, 2009). The measured profiles from radiosonde and RWPs are mostly available for DA during the 1-day assimilation period with relatively few instances of missing data. Unfortunately, the RL had a certain fraction of data eliminated by quality control on 29 and 30 August. The data filtering for RL observations is done as follows. First, biases of retrieved temperature and specific humidity are calculated on a level-by-level basis based on linear interpolation of radiosonde data in time. Then, the normalized bias ratios are acquired by taking the biases divided by the lidar observed value. With the bias ratio given for each data point, data filtering can be carried out by eliminating data when its bias ratio exceeds a certain threshold (0.05 in the current study). We found the data quality issue to be more serious for temperature than specific humidity; therefore, retrieved temperature is excluded from the assimilation data set.

The locations of ARM's surface sites are marked by magenta crosses in Figure 1. In addition, the measurements of Oklahoma Mesonet (McPherson et al., 2007), displayed as empty triangles in Figure 1 and providing statewide coverage of meteorological conditions (horizontal wind components, pressure, temperature,

relative humidity), are also assimilated to assist constraint of surface conditions in a wider region. We note that most of the ARM measurements are not assimilated by operational weather forecast systems since they are not available in real time. Therefore, our results also demonstrate the potential value of including additional regional-scale observations on model predictions.

In GSI, the observation errors for types of measurements are height dependent and read in from a table. Except for RL, the observation errors of other assimilated observations are adopted from an error table that is defined and used for the NAM. In order to construct moisture observation error of RL, we compute the mean deviation of mixing ratio (g kg^{-1}) between Central Facility radiosonde and filtered RL from surface to 3-km mean sea level (MSL) over the whole period of second IOP (not shown). It indicates that the deviation is about 0.4 g kg^{-1} with slight variation in height. Since the deviation is relatively small in comparison with the actual mixing ratio value (~10 to ~20 g kg^{-1}), we apply the moisture observation errors of radiosonde to the case of RL.

3. Case Description and Experimental Design

3.1. Meteorological Conditions on 30 August 2016

Several cases with shallow-to-deep cloud transitions were observed during HI-SCALE (Fast et al., 2018). Most of these cases occurred during the summer IOP because fair-weather conditions and stronger surface heating are more common in summer than in spring. Among those events, 30 August of 2016 is selected for this study because of the widespread formation of ShCu, which was strongly influenced by land-atmosphere coupling as described by Fast et al. (2019). During the morning, as shown in Figure 2a, a band of high clouds extended along a trough from eastern New Mexico toward the northeast into western Kansas in association with a slow-moving frontal system over the central United States. This cloud band lingered over western Oklahoma for the entire day. Clear-sky conditions were observed over central and eastern Oklahoma during the morning until ShCu formed around 16 UTC (10 Central Standard Time [CST]) (Figure 2b) as the CBL became deeper than the LCL. Widespread cumulus formed over southeastern Kansas, central and eastern Oklahoma, and northeastern Texas by 1815 UTC (Figure 2c). After 19 UTC (Figures 2d–2f), ShCu at some locations transitioned into deeper convective clouds. We note that the GOES-13 reflectance shown in Figure 2 has a resolution of ~1 km. Consequently, individual ShCu may not be fully represented.

3.2. DA Strategy and Experimental Design

A cycling assimilation strategy is applied to all DA experiments in this study that consists of alternating periods of DA and no DA that brings the three-dimensional atmospheric conditions closer to observations over time. Earlier studies have suggested that the use of a cycling assimilation strategy could aid in generating optimal analyses due to constraints by consecutive data sets while better maintaining model balance (Xiao & Sun, 2007). The current assimilation schedule, depicted in Figure 3, for our default DA experiment (named "4DEnVar" hereafter) begins with a 12-h spin-up period initialized by NCEP FNL reanalysis to generate finer-scale meteorological conditions. Following that, a 24-h cycling assimilation period is then performed in which five individual assimilations are conducted every 6 h with the identical DA configuration to assimilate observations collected in each time interval. In each assimilation, a 3-h 4DEnVar window is designed to assimilate valid observations at three bins which correspond to three hourly updated backgrounds. The model fields including zonal and meridional winds, specific humidity, temperature, and pressure are updated based on analyzed increments and then used as the input file for the reinitialization of subsequent forecast. The NCEP GDAS observations are assimilated in all four domains, while the additional Oklahoma Mesonet and ARM SGP observations are only assimilated in domains 3 and 4. Once the whole procedure is completed, the last analysis of DA at 12 UTC of 30 August is used to initialize regional cloud simulations over all domains for a 12-h forecast period. Note that the localization parameters such as horizontal and vertical localization distances that are used in generating both static and ensemble-estimated background error covariances are adjusted accordingly with domain resolutions as they are employed in GSI to define the appropriate influential range of analysis increment. For instance, the horizontal localization distances for ensemble-estimated background error covariance are set as 110, 50, 30, and 10 km for domain 1, 2, 3, and 4, respectively. The vertical localization distance is kept at 3-sigma levels for all four domains.
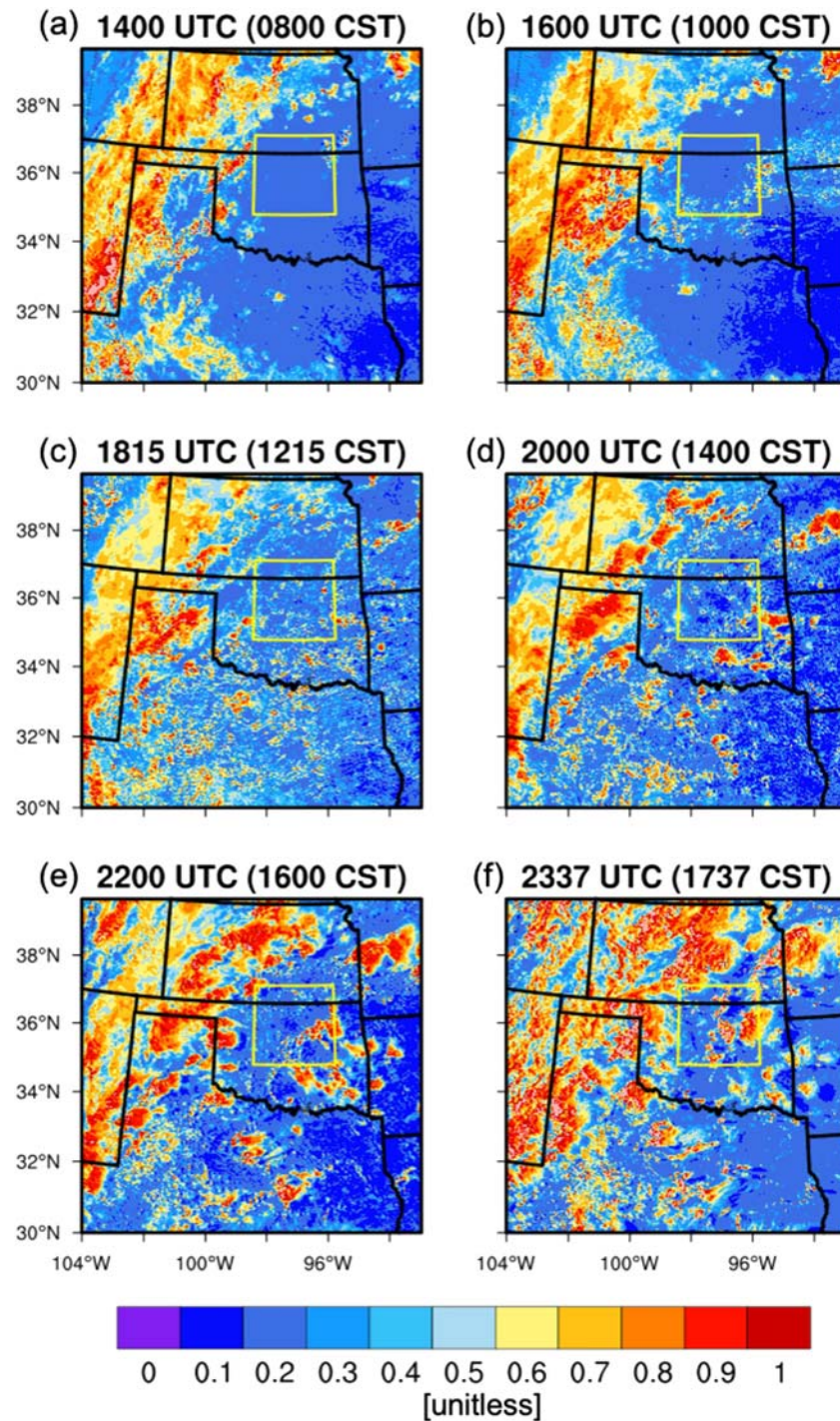
Figure 2. Snapshots of observed GOES-13 visible band reflectance at (a) 1400, (b) 1600, (c) 1815, (d) 2000, (e) 2200, and (f) 2337 UTC on 30 August 2016 over the Southern Great Plains. The rectangle indicates where domain 4 is located.

Three additional experiments are designed to better understand the influences of DA in different aspects. First, a pure WRF forecast experiment initialized at 00 UTC of 30 August by using NCEP FNL reanalysis (named "FNL_BC") is performed as it reflects what will be obtained if there is no assimilation involved. Second, we conduct a WRF forecast initialized at 12 UTC by employing 3-km resolution analyses of National Oceanic and Atmospheric Administration (NOAA)'s High-Resolution Rapid Refresh (HRRR) product (Benjamin et al., 2016) since it is the highest-resolution reanalysis available to initialize WRF (named
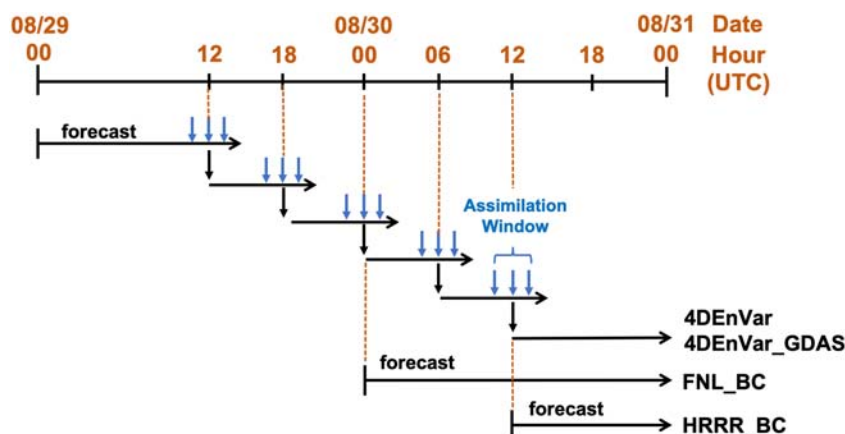
Figure 3. Schematic diagram to illustrate workflows including assimilation strategy for all the experiments. The 4DEnVar and 4DEnVar_GDAS experiments share the same 4DEnVar hybrid assimilation strategy but differ in assimilated observations. The three arrows represent 3 h in each assimilation window. FNL_BC and HRRR_BC are two pure forecast experiments which are initialized by different reanalyses at 00 UTC and 12 UTC of 30 August, respectively.

"HRRR_BC" experiment). Note that in HRRR_BC, domain 1 is not used since HRRR analysis domain does not cover all of domain 1. The third experiment is designed to examine the data impact coming from non-ARM sources other than operational data stream (e.g., satellite observations). Therefore, a sensitivity DA experiment named "4DEnVar_GDAS" is carried out in which only observations from NCEP GDAS data bundles are assimilated. Brief descriptions for each aforementioned experiment are summarized in Table 2.

4. Results

4.1. Simulated Clouds Over Oklahoma and Adjacent Regions

Since the current DA experiments do not update hydrometeors, the variation of predicted clouds in all the experiments can be solely attributed to changes of meteorological conditions between those experiments. Since the DA period ends at 12 UTC, the following discussion will be focused on how the clouds evolve during daytime of 30 August, from 12 UTC to 00 UTC 31 August, from different experiments. Domain 3 covers all of Oklahoma and southern Kansas with a grid spacing of 4 km. The cumulus parameterization is turned off in domains 3 and 4, since it is not necessarily valid at those spatial scales. While shallow convective clouds are predicted by the model by the microphysics parameterization, we note that the grid spacings of 4 and 1.33 km only resolve ShCu cloud populations.

Table 2
List of All Experiments Including Simulations With and Without Coupling With Data Assimilation System

		Assimilated data set	
Experiment	I.C. & B.C. (resolution)	NCEP GDAS	ARM and Oklahoma Mesonet
FNL_BC	NCEP FNL (1°)	N	N
HRRR_BC	HRRR (3 km)	N	N
4DEnVar_GDAS	NCEP FNL (1°)	Y	N
4DEnVar	NCEP FNL (1°)	Y	Y

Note. The assimilated data sets in each experiment are also indicated. Note that the NCEP GDAS data set includes both conventional (radiosonde, aircraft, ship, buoy, wind profiler, etc.) and satellite data (daily AMSU-A brightness temperature, HIRS-4 radiance, Microwave Humidity Sounder [MHS] brightness temperature), while the ARM data set includes observations from radiosonde, radar wind profiler, Raman lidar, and Mesonet.
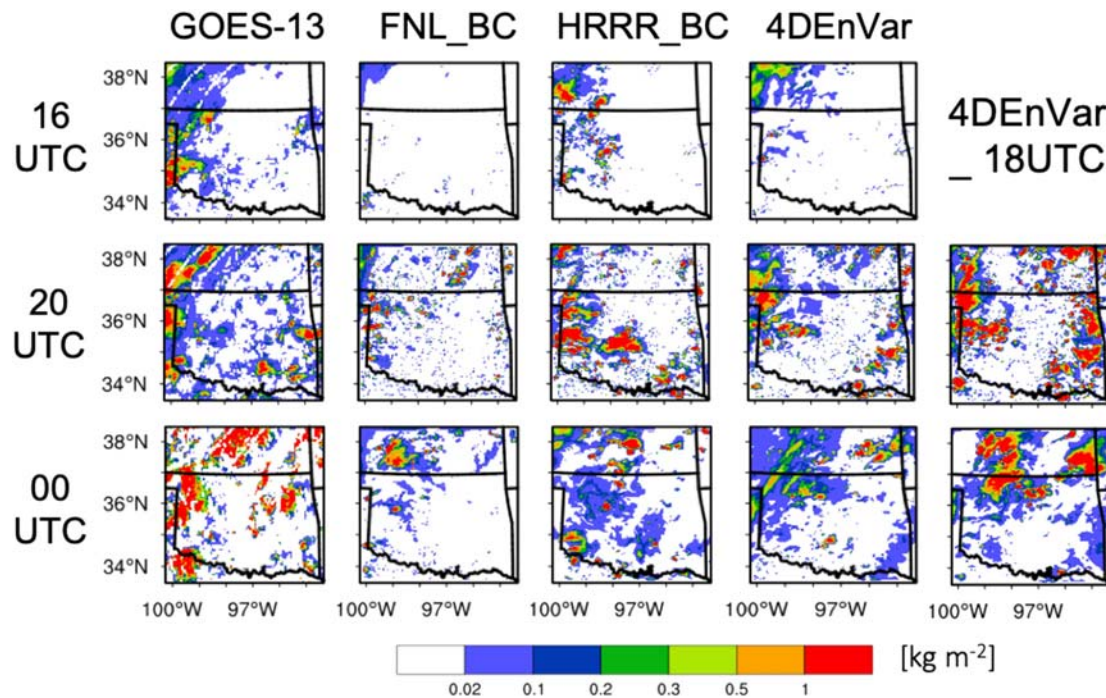
Here, we use metrics of total (liquid and ice) cloud water path (CWP) and liquid water path (LWP) to directly compare with gridded ($\Delta x = 5$ km) retrieved products of the GOES-13 satellite measurements (Minnis et al., 2008, 2011). The Visible Infrared Solar-Infrared Split Window (VISST) algorithm used for retrieval is not only able to estimate the amount of cloud water within each vertical grid column, but also to identify the cloud top height and phase of each grid. The simulated CWP and LWP of each experiment are obtained for each column within domain 3 via equations

$$\text{CWP} = \int_{p_t}^{p_s} (q_l + q_s + q_i) g^{-1} dp, \tag{1}$$

$$\text{LWP} = \int_{p_t}^{p_s} q_l g^{-1} dp. \tag{2}$$

where q_l, q_s, and q_i represent liquid, snow, and ice cloud mixing ratios (kg kg^{-1}), respectively. G is gravitational acceleration (m s^{-2}), and dp is the pressure increment (Pa) between two layers in the atmosphere, with p_s and p_t representing the surface and model top pressures, respectively.

Figure 4. Snapshots at 16 and 20 UTC of 30 August and 00 UTC of 31 August 2016 that display cloud water path from GOES-13 retrievals and simulated results of experiments FNL_BC, HRRR_BC, and 4DEnVar. 4DEnVar_18UTC represents forecast result with reinitialization at 18 UTC. Note that all the values are shown by shading in unit of kg m^{-2}.

Figure 4 displays plots of the GOES-13 observed CWP along with simulated CWP from experiments FNL_BC, HRRR_BC, and 4DEnVar at three selected times (16, 20, and 00 UTC). Note that 4DEnVar_GDAS results are not included in Figure 4 since they are similar to those from 4DEnVar. The observed CWP shows the presence of a cloudy region over the western domain at 16 UTC, which is part of the band of clouds extending from western Texas into Nebraska as shown in Figure 2. At 20 UTC, the total cloud fraction (CF) increases, and some of the clouds over the western and southeastern sides of the domain transition into more convective cloud populations as the amounts of cloud water significantly increase. At 00 UTC, most of the clouds in the domain have become even deeper and propagate toward the central Oklahoma from both the west and east sides of the domain.

CWPs from experiments FNL_BC, HRRR_BC, and 4DEnVar exhibit diverse cloud development within domain 3, indicating that the uncertainty of large-scale and mesoscale ambient environmental conditions is fairly large and significantly impacts the formation and location of cloud populations. Without additional DA, FNL_BC has difficulty in reproducing clouds over the western domain at 16 UTC, which implies that aspects of the large-scale trough system are poorly predicted. In contrast, the HRRR_BC better represents the presence of clouds in this region due to the fact that hydrometeors are also updated in HRRR analysis in addition to other conventional DA. 4DEnVar is able to recover some of the clouds over western domain through cycled assimilation but is relatively weaker than HRRR_BC since the hydrometeors are not updated in current DA configuration. After 4 h of integration at 20 UTC, FNL_BC has more clouds simulated, but it misses the deeper clouds observed over southeastern Oklahoma. HRRR_BC predicts more cloudiness comparable to observations but produces an area of spurious convection over south central Oklahoma. 4DenVar overall exhibits better qualitative agreement with the observed cloud distributions with the deepest convection over western and southeastern sides of the domain. Compared with observed CWP, all experiments in general simulate shallower convections in the late afternoon as shown in snapshots at 00 UTC.

To quantitatively evaluate the CFs simulated in the experiment, we compute the total CF and liquid CFs (LCFs) based on the fractions of columns within domain 3 in which CWP and LWP are greater than 1 g m^{-2}. Similarly, the observed CF and LCF are obtained by calculating the fraction of pixels with
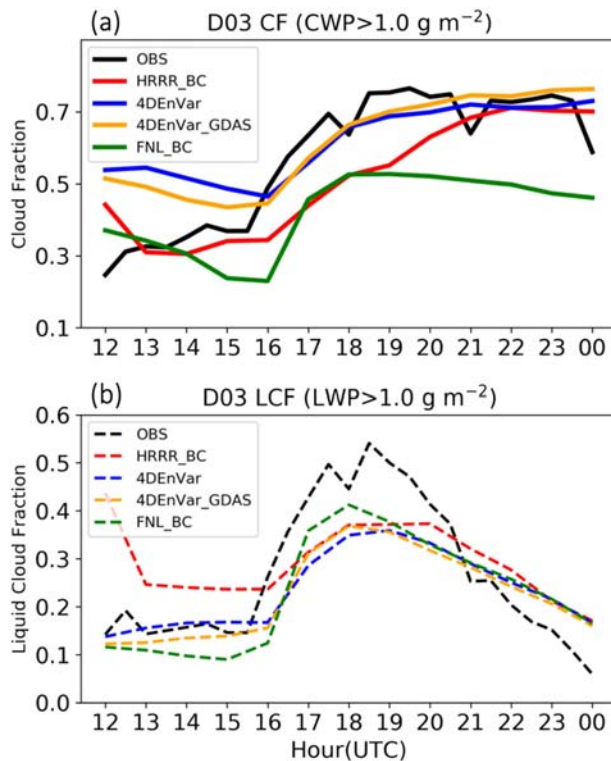
Figure 5. Comparison of domain 3 simulated (a) total cloud (including both liquid and ice) fraction (CF) and (b) liquid cloud fraction (LCF) against corresponding observations (GOES-13 gridded retrievals) in time series. The colored lines represent results calculated from observation and different experiments as shown in the legends.

GOES-13 satellite radiance retrieved CWP and LWP greater than 1 g m^{-2} within the same domain. Since there are mixed-phase pixels identified, the LCF accounts for pixels identified as in the liquid phase as well as additional mixed-phase (liquid–ice, liquid–clear-sky) pixels in the calculation. Note that the grids with retrieved cloud top heights higher than 4.5-km MSL (approximated melting level) have been excluded in the calculations of observed LCF to further filter out the non-liquid phase cloud pixels. In addition, simulated CWP and LWP are interpolated onto the GOES-13 retrieval grids to avoid possible issue induced by different total grid numbers in CF and LCF calculations.

Figure 5a depicts the evolution of observed and simulated CFs over the whole domain 3 during daytime of 30 August from 12 to 00 UTC. Satellite measurements indicate a rapid increase in CF from ~0.49 to ~0.75 between 16 to 19 UTC and then CF remained relatively steady the rest of the afternoon. Among the simulated CFs, 4DEnVar (blue line) and 4DEnVar_GDAS (orange line) are in better agreement with the observed curve, although the increase in cloud coverage is slower than observed. FNL_BC and HRRR_BC both capture the signal of rapid increase in cloud coverage from 16 to 18 UTC, but CFs are much lower than observed. FNL_BC is 0.2 to 0.3 lower than the observations, which is consistent with Figure 4.

To assess the model predictions of shallow convective cloud populations, liquid cloud water fraction over the domain as a function of time is shown in Figure 5b. A similar but more prominent signal of rapid cloud formation and growth occurs between 16 and 18 UTC, demonstrating that shallow convection is responsible for the rapid increase in both CF and LCF. Even though LCFs among the experiments vary before 16 UTC, all the simulations qualitatively reproduce the life cycle of liquid clouds, including a peak around noon and subsequent decay in the afternoon.

We also evaluate precipitation that is produced as a fraction of the shallow convection transition to deeper convection. The 12-h accumulated precipitation (12 UTC of 30 August to 00 UTC of 31 August) over domain 3 from all the experiments are shown in Figure 6. The NCEP Stage IV multisensor (radar and gauges) precipitation analysis (4-km resolution) is introduced here as the observation (Lin & Mitchell, 2005). Figure 6a shows patches of accumulated rainfall over 30 mm that were mainly observed over the western part of domain 3, which is associated with convection along the synoptic-scale trough. More scattered and lighter precipitation is found over the eastern side of domain and can be attributed to isolated and transient deep convection. In contrast, central Oklahoma is mostly precipitation free during the 12-h period. Compared with the Stage IV precipitation analysis, FNL_BC produces much less precipitation (Figure 6b) in general. HRRR_BC produces patches of precipitation along the trough near the western boundary of the domain more comparable to observed amount of precipitation, but it produces too much precipitation over south central Oklahoma and not enough precipitation over northeastern Oklahoma (Figure 6c). 4DEnVar_GDAS produces light and scattered precipitation over much of the domain, having less agreement with the observation (Figure 6d). While the CFs from 4DEnVar and 4DEnVar_GDAS are similar (Figure 5b), the rainfall pattern from 4DEnVar is better when compared with the observed pattern (Figure 6e).

Simulated hourly mean rain rates over domain 3 during the 12-h period mentioned earlier are also compared to NCEP Stage IV hourly data and given in Figure 7. The observed curve indicates a small rain rate peak near 14 to 15 UTC, which is only captured by HRRR_BC. After 18 UTC (local noon), the observed rain rate increases in time during the rest of the afternoon, reflecting isolated areas of deep convection over domain 3. This trend, however, is not reproduced by any of the model experiments. Instead, all the experiments produce a rainfall peak around 19 to 20 UTC. The use of DA in 4DEnVar and 4DEnVar_GDAS produces a higher rain rate than FNL_BC. The 4DEnVar_18UTC experiment shows a significant increase in rainfall after 19 UTC, but the highest rain rates occur earlier than observed.
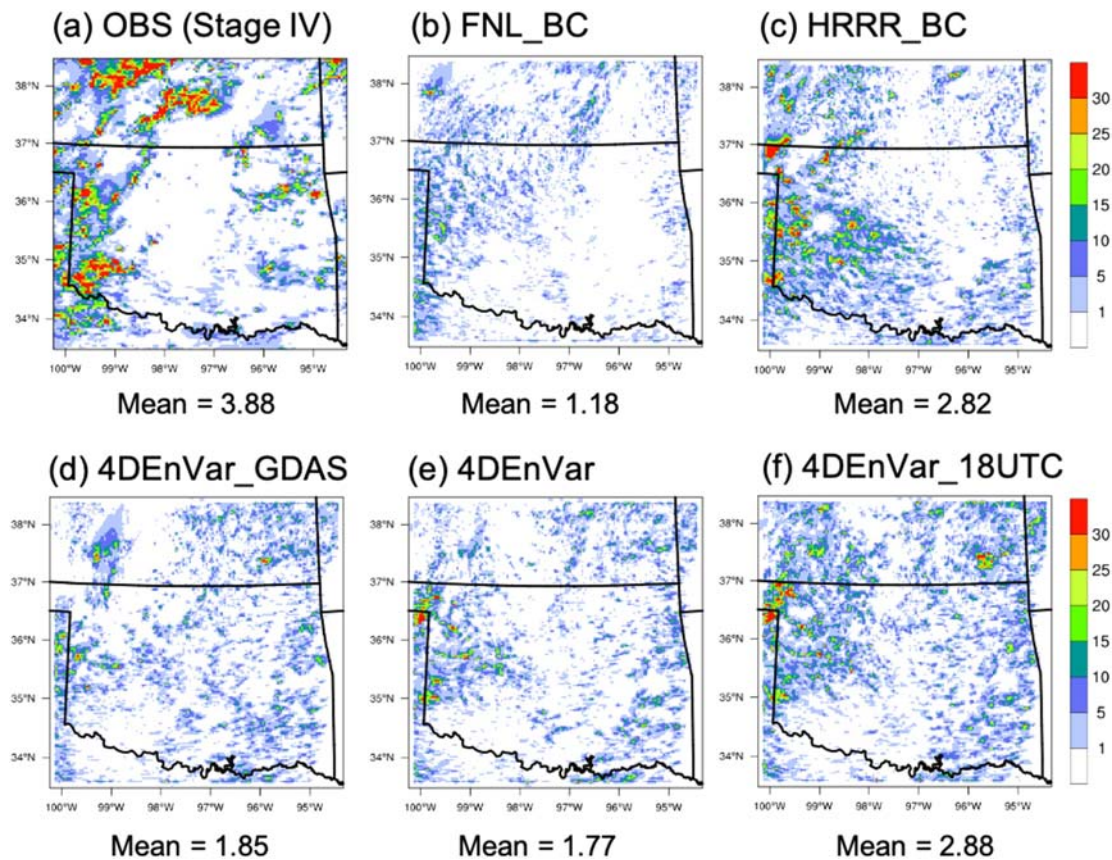
Figure 6. Twelve-hour (from 12 UTC 30 August to 00 UTC 31 August) accumulated rainfall amounts (mm) over domain 3, including results of (a) Observation (NCEP Stage IV), (b) FNL_BC, (c) HRRR_BC, (d) 4DEnVar_GDAS, (e) 4DEnVar, and (f) 4DEnVar_18UTC. The mean values (mm) for each data are listed under each plot.

4.2. Simulated Clouds Near the ARM SGP Site

Various data sets on ShCu and BL properties are utilized here to comprehensively evaluate simulated cloud fields and the corresponding atmospheric conditions near the SGP Central Facility; therefore, we now focus on a more localized area corresponding to the innermost domain (Δx = ~1.33 km). The impact of resolution is shown in Figures 8a and 8b, where CWP simulated over domain 3 from 4DEnVar at 20 UTC is zoomed in to the same area as domain 4. Even though the general cloud patterns are similar between the two domains, the size of shallow convective clouds is smaller and their number tends to be larger in higher-resolution simulation than in lower-resolution one as expected. The gridded CWP retrievals from measurements of MODIS on the NASA Earth Observing System Aqua satellite at 1950 UTC (1-km grid spacing, Figure 8c) and GOES-13 at 1800 UTC (5-km grid spacing, Figure 8d) are introduced as the observations. A comparison of Figures 8c and 8d indicates that GOES-13, with a resolution of 5 km, captures the larger-scale spatial distribution of cloud cover but smooths the smaller and discrete shallow cloud clusters (as seen in the MODIS retrieval) by filling in the clear-sky areas between individual clouds. As a result, the GOES-13 retrieval has a higher CF of ~0.6 than the number computed from MODIS retrieval (~0.4), implying that a larger bias could exist during periods with more shallow subgrid clouds. Domain 4 of 4DEnVar (Figure 8b) simulates a relatively wider area of the high clouds over the northwestern side than the observed MODIS CWP (Figure 8c). In addition, the simulated clouds over the eastern part of domain are less aggregated than those in the MODIS retrieval. We also notice that the
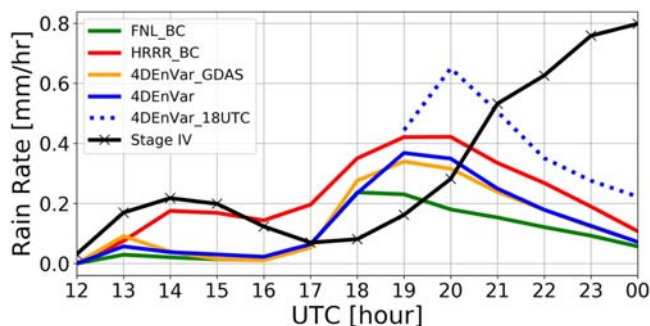

Figure 7. Comparison of simulated mean hourly rain rate (mm h^{-1}) against NCEP Stage IV observation over the whole domain 3 area from 12 UTC of 30 August to 00 UTC of 31 August. Colored lines denote results calculated from observation and different experiments as shown in the legends.
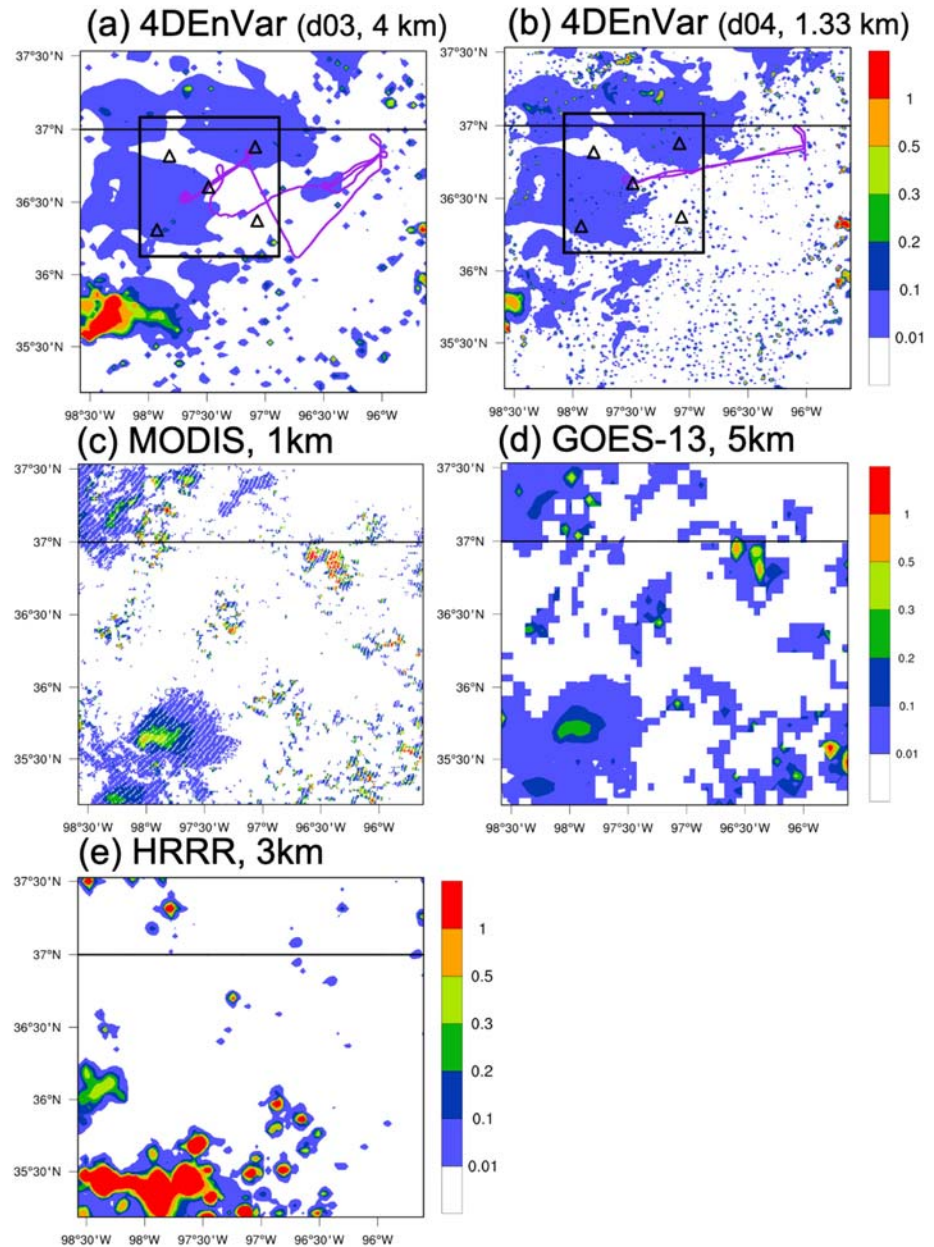
Figure 8. Snapshots of cloud water path (shading, kg m^{-2}) at 20 UTC of 30 August, including results from (a) zoomed-in 4DEnVar domain 3 simulation ($\Delta x = 4$ km), (b) 4DEnVar domain 4 simulation ($\Delta x = 1.33$ km), (c) MODIS Aqua retrieval (1-km resolution), (d) GOES-13 retrieval (5-km resolution), and (e) HRRR operational forecast ($\Delta x = 3$ km). In (a) and (b), the locations of Doppler lidars are depicted by black triangles. The black square represents the localized area used in area-mean calculations for simulations, and the purple lines in (a) and (b) represent the tracks of G-1 flight in morning and afternoon missions, respectively.

operational forecast from HRRR (version 2 in 2016), shown in Figure 8e, produces mostly clear skies over the SGP site where shallow clouds are observed. This is consistent with an issue in version 2 of the HRRR system noted by Benjamin et al. (2016), namely, an underprediction of resolved clouds, especially ShCu, which can lead to an overestimation of the shortwave solar flux at the surface. More recent versions of HRRR now represent subgrid-scale clouds using the MYNN eddy diffusivity mass flux (EDMF) boundary layer parameterization (Olson et al., 2019) that would likely better estimate fractional cloudiness for the conditions observed on 30 August.
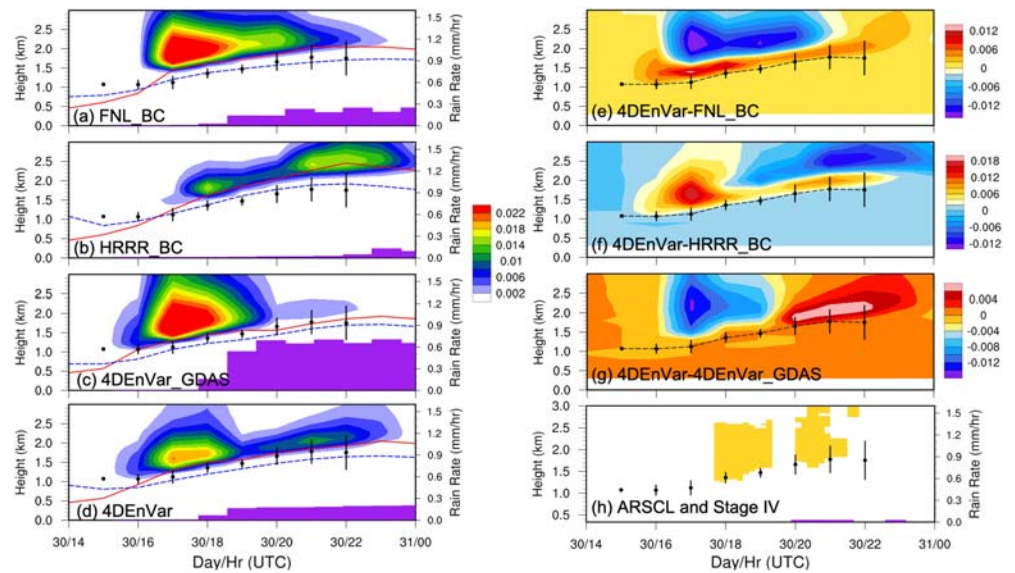
Figure 9. Time-height comparison of domain 4 simulated area-mean (see the black square area depicted in Figure 8a) liquid water mixing ratio (shading, kg kg^{-1}) from (a) FNL_BC, (b) HRRR_BC, (c) 4DEnVar_GDAS, and (d) 4DEnVar experiments below 3 km over a period from 14 UTC 30 August to 00 UTC 31 August. In each subplot, the corresponding convective boundary layer height and lifting condensation level are overlaid with solid line in red and dashed line in blue, respectively. The differences between 4DEnVar and other three experiments are given in (e) 4DEnVar and FNL_BC, (f) 4DEnVar and HRRR_BC, and (g) 4DEnVar and 4DEnVar_GDAS. The ARSCL cloud mask is illustrated in yellow shading in (f). The hourly mean and standard deviation of estimated cloud base heights from Doppler lidar measurements are marked by black dots and error bars. The area-mean hourly rain rate (mm h^{-1}) of each simulation (a–d) and corresponding Stage IV rain rate observation (h) are overlaid in purple bars.

In addition to the spatial distribution of clouds, it is important to evaluate the simulated cloud base height (CBH) when compared with estimates obtained from the network of five DLs (shown in Figures 8a and 8b). Simulated area-mean cloud water mixing ratio (shown by color shading in Figure 9) is computed below an altitude of 2.5 km within a 133 × 133 km area centered at Central Facility which is slightly larger than the area of DL network. Moreover, simulated time series CBL heights (identified by maximum gradient of virtual potential temperature profiles) and LCL averaged over the same area are overlaid with red solid and blue dashed lines. The CBH estimated by DLs are marked by black dots with error bars showing the mean and standard deviation at each hour. The observed CBH indicates that the initiation of shallow convective clouds in the region occurs nearly at 15 UTC (09 CST) at an altitude of ~1.1 km. Then, CBH rises in time and reaches the maximum altitude of ~1.7 km. The error bars also show that spatial variability in CBH increases during the late afternoon. The initiation time of shallow convective clouds is delayed about 1 and 2 h in FNL_BC and HRRR_BC (Figures 9a and 9b), respectively, while the initiation time from both 4DenVar_GDAS and 4DEnVar are closer to the observed time (Figures 9c and 9d). In the meantime, the simulated LCLs in both 4DEnVar_GDAS and 4DEnVar are relatively lower than FNL_BC and HRRR_BC. To illustrate differences in the simulated cloud base and depth, the results from FNL_BC, HRRR_BC, and 4DEnVar_GDAS are subtracted from 4DEnVar. Figures 9e–9g clearly show that the cloud bases simulated by both FNL_BC and HRRR_BC are on average ~500 m higher than observed (Figures 9e and 9f). While the cloud base from 4DEnVar_GDAS and 4DEnVar are very similar (Figure 9g), the average cloud liquid water mixing ratio from 4DEnVar is lower during the morning and higher during the afternoon. The horizontal cloud distributions (not shown) also indicate that 4DEnVar_GDAS has fewer clouds near the SGP site after 19 UTC.

The cloud mask product provided by Active Remote Sensing of Clouds (ARSCL; Clothiaux et al., 2001) is given in Figure 9h to illustrate the observed cloud base and depth at the Central Facility site. It combines measurements from remote sensors in Central Facility to produce an objective determination of hydrometeor height distributions. It indicates that ShCu was first detected just before noon (18 UTC) at the Central Facility site and suggests that the estimation of CBHs is quite robust as both observational data
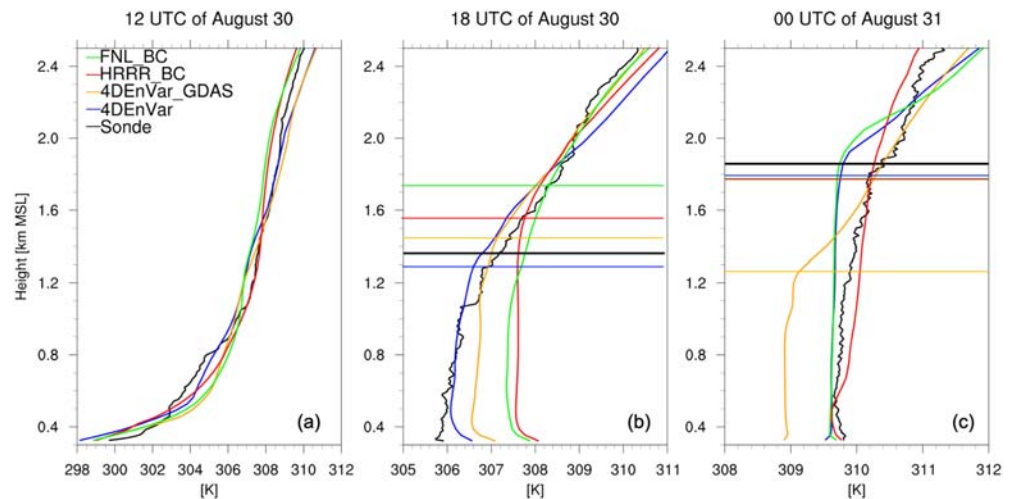
Figure 10. Comparison of simulated virtual potential temperature (K) profiles below 2.5 km MSL with the radiosonde observations at the ARM SGP Central Facility at (a) 12 UTC, (b) 18 UTC of 30 August, and (c) 00 UTC of 31 August. Profiles of simulations are shown in lines with different colors indicated in the legend. In (b) and (c), the horizontal lines denote the convective boundary layer heights estimated from corresponding virtual potential temperature profiles.

sets show similar results. The liquid cloud depth identified by ARSCL cloud mask varies from ~1.2 to over 1.5 km, suggesting that simulated shallow convective cloud depths are also comparable with observations. Nevertheless, it should be noted that the ARSCL product is not necessarily representative over an area as large as domain 4.

To examine the capability of model in reproducing the observed diurnal cycle of precipitation near the ARM SGP site, the simulated area-mean (same domain as the black square depicted in Figures 8a and 8b) hourly rain rate of each model experiment is depicted in Figure 9. The areal mean of hourly rain rate from NCEP Stage IV precipitation analysis is introduced as a reference in Figure 9h. While 4DEnVar_GDAS (Figure 9c) simulates relatively higher rain rate than other experiments, overall, it indicates that all the model experiments produce light and persistent precipitation during the presence of ShCu populations in the afternoon (after 18 UTC). Although all simulated rain rates are higher than what is observed in the Stage IV data, it should be noted that the comparison here is a qualitative evaluation as the 4-km Stage IV data set may have larger uncertainty in situations of lower rain rate.

4.3. Examination on Environmental Conditions Near the ARM SGP Site

In this section, the simulations are evaluated using observations obtained at or near the ARM SGP site to provide insights into the changes of meteorological conditions that are made by DA and how those changes then alter the evolution of shallow convective clouds.

4.3.1. Radiosonde

The evolution of the CBL strongly controls formation and growth of shallow convection; therefore, the observed and simulated virtual potential temperature (θ_v) profiles below 2.5-km MSL are shown at 6-h intervals in Figure 10. Note that the simulated profiles from all experiments are obtained by averaging over nine horizontal grid points from domain 4 closest to the Central Facility. The simulated CBL heights at 18 and 00 UTC are then identified as the level with largest gradient in the mean virtual potential temperature profile for each experiment, while the observed CBL heights are estimated by the method provided in Liu and Liang (2010). All the CBL heights are denoted by overlaid horizontal lines. During the early morning at 12 UTC (06 CST), all of the model experiments are similar and they reasonably reproduce the observed radiative cooling near the ground (Figure 10a). However, the simulated profiles at noon (18 UTC) in Figure 10b are more diverse. Predictions of θ_v in the CBL differ by as much as 1.6 K. θ_v from 4DenVar is the closest to the radiosonde within the CBL, although 4DEnVar produces a depth of 1.29 km that is about 100 m lower than observed (1.38 km). The other three model experiments have larger positive biases within the CBL; as a result, the mixed layer depths (1.45, 1.56, and 1.73 km for 4DEnVar_GDAS, HRRR_BC, and FNL_BC, respectively) are greater than the observed depth of 1.3 km. The θ_v profiles at 00 UTC
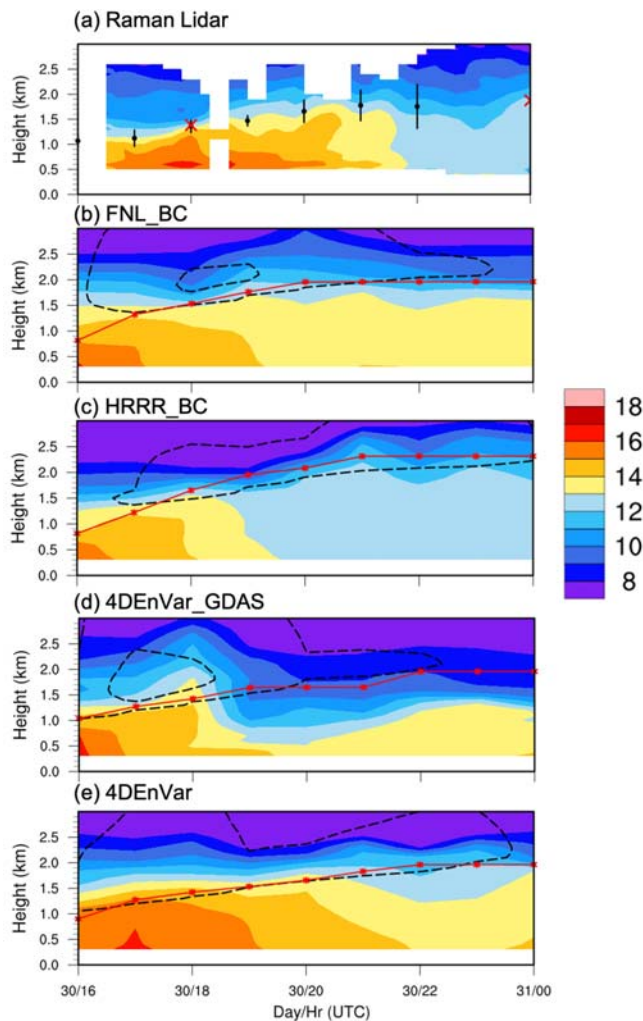
Figure 11. Time-height comparison of specific humidity (shading, g kg^{-1}) at Central Facility below 3 km height over a period from 16 UTC of 30 August to 00 UTC of 31 August, including (a) observation of Raman lidar and simulated results in experiments (b) FNL_BC, (c) HRRR_BC, (d) 4DEnVar_GDAS, and (e) 4DEnVar. In (a), the Doppler lidar estimated cloud base heights and the Central Facility radiosonde estimated convective boundary layer heights are overlaid by black dots (with error bars) and red crosses, respectively. In (b–e), corresponding cloud mixing ratios and convective boundary layer heights are illustrated by black dashed contours (0.002 and 0.2 g kg^{-1}) and red lines with star signs.

(Figure 10c) are even more diverse due in part to differences in whether the simulations produce local deeper, precipitating clouds near the Central Facility site. Nevertheless, three of the experiments simulate CBL depths around 1.8 km which are comparable with observed depth of 1.88 km.

4.3.2. Raman Lidar

The RL deployed at ARM SGP Central Facility retrieves both temperature and specific humidity profiles with high temporal resolution (10 min). As mentioned earlier in section 2.2, the observed moisture profiles have been quality controlled and are valid only after 16 UTC (10 CST) on 30 August. Similar to the comparison with radiosondes, the simulated specific humidity profiles below 3-km MSL are obtained with averaging over nine grid points from domain 4 closest to the Central Facility. The observed vertical variations in specific humidity from the RL along the simulated variations are presented in Figure 11. The observed contour of 13 to 14 g kg^{-1} (yellow shading) moves upward in time from approximately 1 to 1.7 km due to CBL growth (Figure 11a), which matches quite well with the evolution of DL estimated CBHs denoted by black dots. The CBL heights estimated by Central Facility radiosonde profiles at 18 and 00 UTC are also given in Figure 11a. It shows good consistency at 18 UTC and provides an additional guidance of CBL height at 00 UTC when estimates from the DLs were invalid.

In Figures 11b–11e, the simulated moisture variation is overlaid by area-mean cloud mixing ratio as well as the CBL height for each model experiment (same values shown in Figures 9a–9d). FNL_BC, HRRR_BC, and 4DEnVar_GDAS show that contours of 14 to 15 g kg^{-1} are absent after ~19 UTC (Figures 11b–11d), indicating that the boundary layer is drier than RL observations. 4DEnVar has higher mixing ratios during the afternoon up to 22 UTC and is more consistent with observations; however, the simulation slightly overpredicts mixing ratios close to 16 UTC and after 22 UTC (Figure 11e). The dry bias is most pronounced in HRRR_BC, which contributes to a more stable environment which results in higher LCL (Figure 9b) and a cloud base that is too high (Figure 11c). 4DEnVar_GDAS (Figure 11d) exhibits an abrupt increase of moisture above the boundary layer around 18 UTC which is associated with the formation of deeper clouds, and then, entrainment of dry airs appears above top of CBL (~1.5 km) after 20 UTC. Both of the variations did not occur in observation. Our results agree with the major finding of Zhang and Klein (2012), in which they investigated factors that control the evolution of ShCu by analyzing 13-year integrated observations measured at the ARM SGP site and found that the relative humidity within boundary (below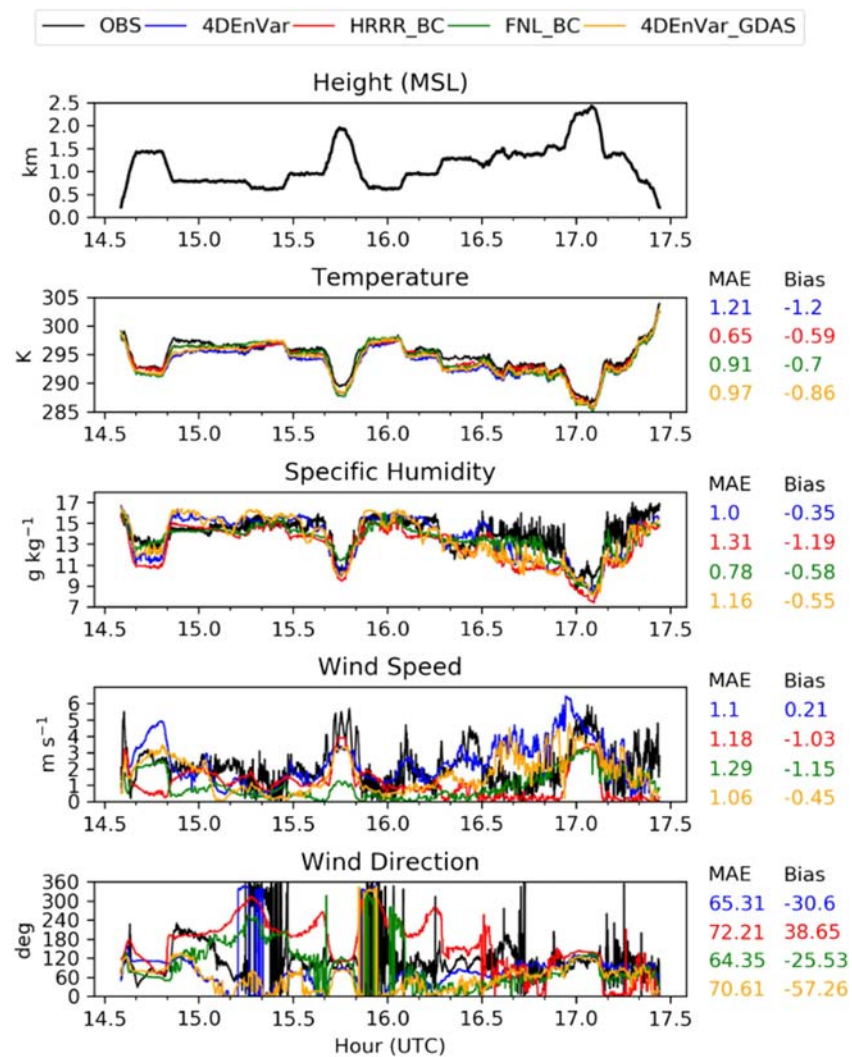 1.5-km MSL) plays the biggest role in modulating ShCu among many other environmental parameters. In summary, it suggests that moisture within CBL is quite sensitive in modeling shallow convective clouds of cloud-system resolving model and is thus essential to be accurately represented.

4.3.3. G-1

On 30 August, the ARM's research G-1 aircraft collected meteorological, trace gas, and aerosol measurements near the ARM SGP site. Two missions were completed just before (1435–1726 UTC) and after (1832–1932 UTC) noon when the LCF reaches its maximum. Exact flight paths for the morning and afternoon mission are depicted in Figures 8a and 8b, respectively. The meteorological observables measured along the flight tracks provide a more regional context for the meteorological conditions at the Central Facility. The model outputs from domain 4 at 1-h interval are interpolated in time and space to match the position of the G-1 aircraft (Fast et al., 2011). Comparisons for the morning mission (Figure 12) show that

Figure 12. Comparison of the observed (black line) and simulated (colored lines) temperature (K), specific humidity (g kg^{-1}), wind speed (m s^{-1}), and wind direction (degree) along the G-1 flight path during the morning of 30 August 2016. The legend at the top denotes the colors used to represent OBS, 4DEnVar, HRRR_BC, FNL_BC, and 4DEnVar_GDAS experiments. The mean absolute error (MAE) and bias for each variable and simulation are listed with corresponding color on the right side of each subplot.

all four model experiments qualitatively represent the spatial and temporal variations in temperature and specific humidity. The mean absolute error (MAE) and bias are provided in the columns at the right side of each subplot, showing the MAE range between 0.65 and 1.21 K for temperature and biases in a range of −0.59 and −1.2 K, in which HRRR_BC has the lowest MAE and bias among experiments but does not differ much when they are scaled by the real temperature. The range of MAE for specific humidity is between 0.78 and 1.31 g kg^{-1}, and 4DenVar has the smallest negative bias (−0.35 g kg^{-1}). Simulated wind speed and direction are more diverse among the experiments than the previous two variables. Relatively larger errors in the wind direction are found in some of experiments when the dominant wind is weak at altitudes near surface (~0.5-km MSL). While both the simulated and observed wind directions are similar along the lowest transects, the simulated winds near 1-km MSL are often northeasterly while the observed winds are easterly. The quantitative evaluation shows that the MAEs are quite comparable among all experiments (1.0 to 1.3 m s^{-1} for wind speed; 64.4° to 76.2° for wind direction); however, 4DEnVar does have better skill in terms of biases of wind speed and direction.

Similar comparisons for the afternoon mission are given in Figure 13. The flight track of afternoon flight mission consists of two primary transects at 1.5- and 2.0-km MSL. While the overall performance in
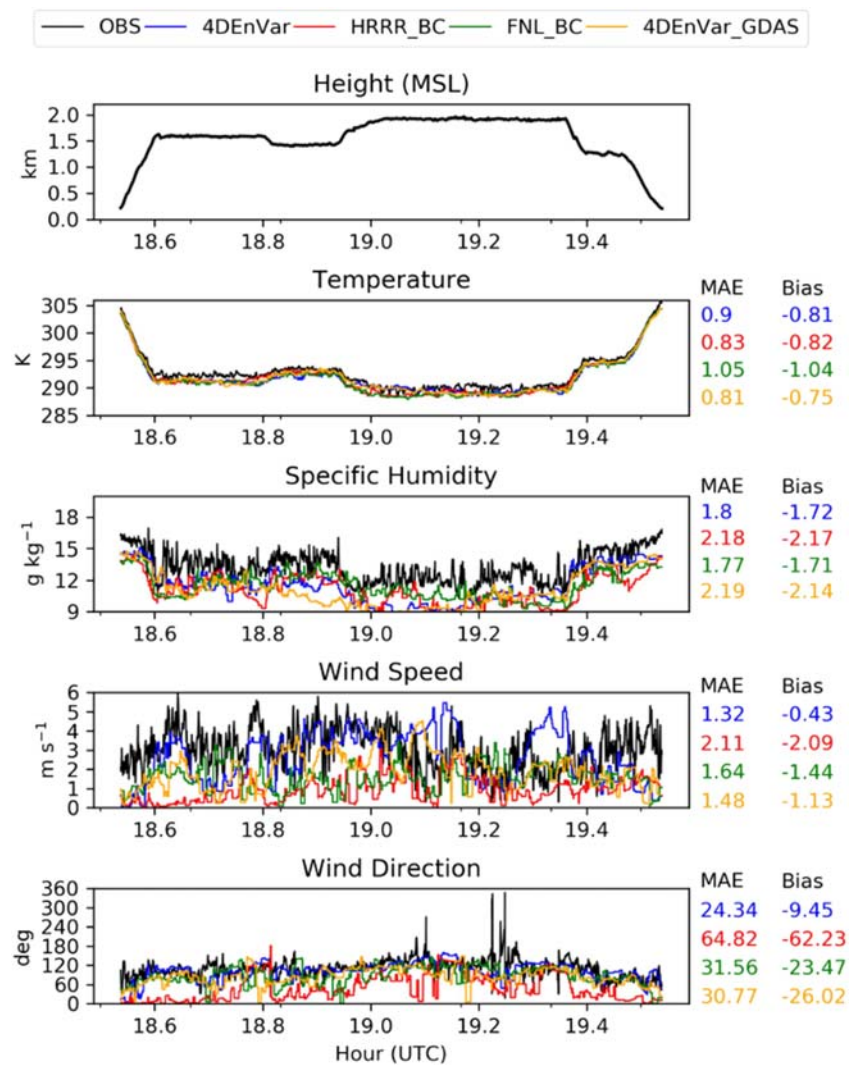
Figure 13. Same as Figure 12, except for the afternoon G-1 flight.

temperatures is similar to the morning flight with MAE ranges from 0.81 to 1.05 K, simulated specific humidity from all the experiments have relatively larger MAEs (between 1.77 and 2.19 g kg^{-1}) and biases (between −1.71 and −2.17 g kg^{-1}) than what are shown previously for morning mission. These results are consistent with the dry bias revealed by the RL profiles (Figure 11). The overall MAEs and biases of wind speed also become slightly larger than the results for morning mission, but the predictions of wind direction become better. It is very clear that the simulated winds from 4DEnVar are better than other experiments at this time as it has the smallest numbers in both MAE and bias.

5. Discussion

5.1. Challenge in Modeling Cloud Transitions: Impact of Additional Assimilation at 18 UTC

The findings in section 4.1 indicate that the intensity of deep convective clouds, including some that are transitioned from shallow convections, is generally underpredicted in the late afternoon in all domain 3 ($\Delta x = 4$ km) simulations. Many components could contribute to this issue, including the relatively coarse resolution (for simulating convective clouds), assumptions in the physics parameterizations, and boundary conditions. One source of error could relate to the time scales of predictability of shallow convective clouds after the last DA cycle (i.e., the forecast lead time). To assess this sensitivity, we conduct an additional 4DEnVar simulation that includes an additional assimilation period at 18 UTC to determine whether further constraining the ambient meteorology has a positive impact on the evolving convective cloud populations.
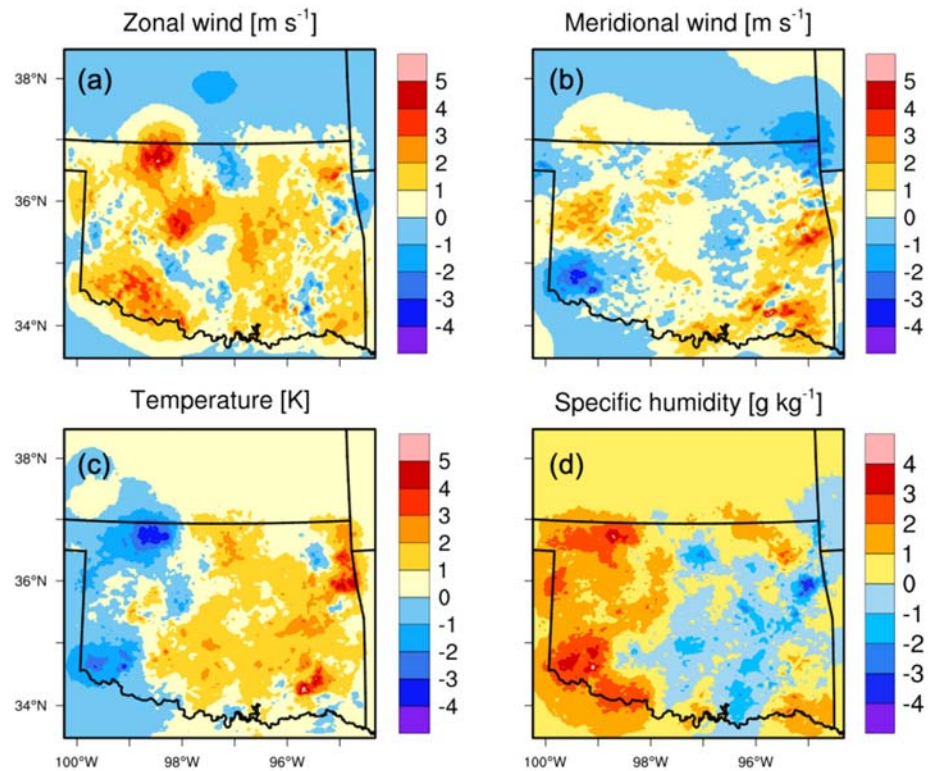
Figure 14. Domain 3 increments of (a) zonal wind (m s^{-1}), (b) meridional wind (m s^{-1}), (c) temperature (K), and (d) specific humidity (g kg^{-1}) at the first model level from the 4DEnVar experiment at 18 UTC on 30 August 2016.

Compared to the forecast initialized at 12 UTC, the new simulation produces more vigorous convection with higher CWP in the updated forecast at 20 and 00 UTC as shown in the column of 4DEnVar_18UTC in Figure 4. As a result, the accumulated precipitation produced by 4DEnVar_18UTC is enhanced (Figure 6f) and becomes more comparable with observations (Figure 6a) as the calculated mean of accumulated precipitation increases from 1.77 to 2.88 mm although the amount remains less than observed value (3.88 mm).

To identify possible mechanisms that strengthen convection during the afternoon after the assimilation period, we examine the analysis increments at 18 UTC at the lowest model level (Figure 14). Relatively large positive zonal wind increments are produced under the eastward-propagated cloud band over the western part of domain (Figure 14a), which corresponds to significant negative temperature and positive specific humidity increments in Figures 14c and 14d. Meanwhile, an area with distinct positive meridional wind and specific humidity increments is produced over the southeastern Oklahoma where deep cloud populations are observed to occur (Figures 14b and 14d). Based on these analyses, the intensification of afternoon convection can be attributed to increased moisture in the vicinity of the convective cells which then enhances the atmospheric instability in subsequent forecast after assimilation. Also, the strengthened gust fronts around the edges of cold pools trigger initiation of new convective clouds that sustains the intensity of the overall convection in the region. To examine this issue, we then track the variation of intensities of cold pools by reviewing simulated and observed surface temperature at 20, 22, and 00 UTC as shown in Figure 15. The results indicate that 4DEnVar_18UTC does simulate stronger cold pools with lower surface temperatures over the western and eastern sides of the domain at 20 and 22 UTC than 4DEnVar. Nevertheless, simulated surface temperatures associated with cold pools are still lower than those observed by the Oklahoma Mesonet.

5.2. Applicability of Current DA Configuration

We recognize that it is beneficial to include a composite analysis of multiple cases to determine whether the current DA configuration robustly improves the simulated diurnal cycle of ShCu. Thus, we applied the
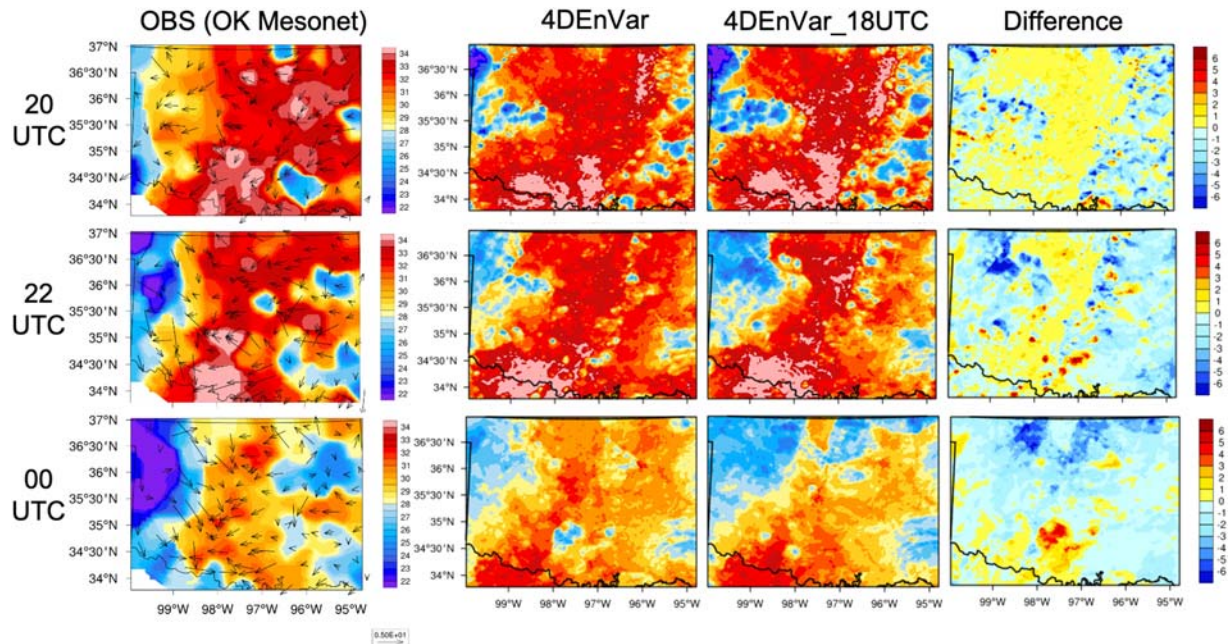
Figure 15. Comparison of observed and simulated surface temperature (°C) over Oklahoma at 20, 22, and 00 UTC. Columns from left to right represent interpolated temperature from the Oklahoma Mesonet overlaid with wind vectors, simulation initialized by 4DEnVar analysis at 12 UTC, simulation initialized by 4DEnVar_18UTC analysis at 18 UTC, and the difference between the two simulations (4DEnVar_18UTC and 4DEnVar).

identical assimilation strategy of the 4DEnVar experiment to the days prior and after 30 August that also had shallow-to-deep transitions in convective clouds over region north-central Oklahoma. In addition to DA experiment, corresponding simulations of FNL_BC and HRRR_BC are also conducted to provide similar comparisons as on 30 August.

To have a consistent evaluation with the case of 30 August, we computed CF and LCF over domain 3 for these 2 days similar to Figure 5. The results for 29 August (Figures 16a and 16b) show that the three simulations underestimate the CF of both total CF and liquid cloud (LCF). While FNL_BC and 4DEnVar have a
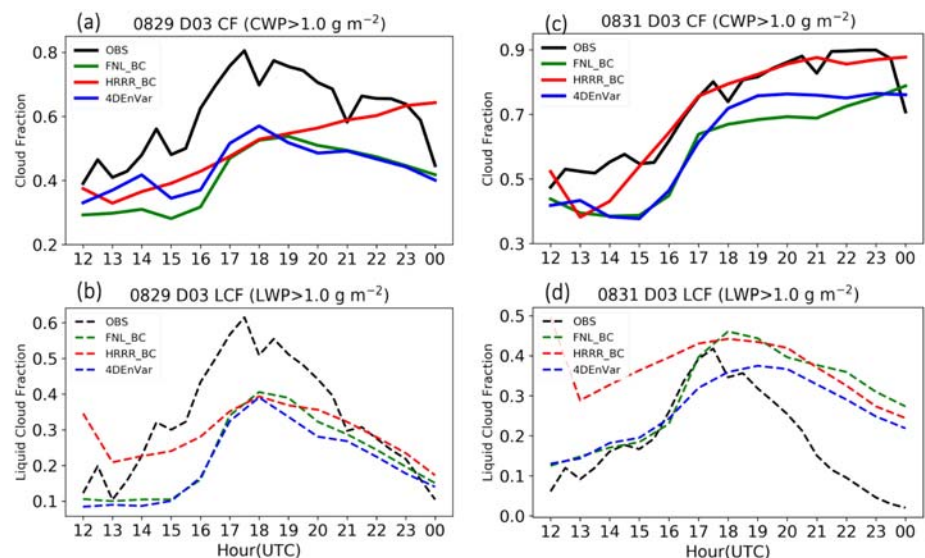

Figure 16. Same as Figure 5 but for the cases of 29 (a and b) and 31 (c and d) August.
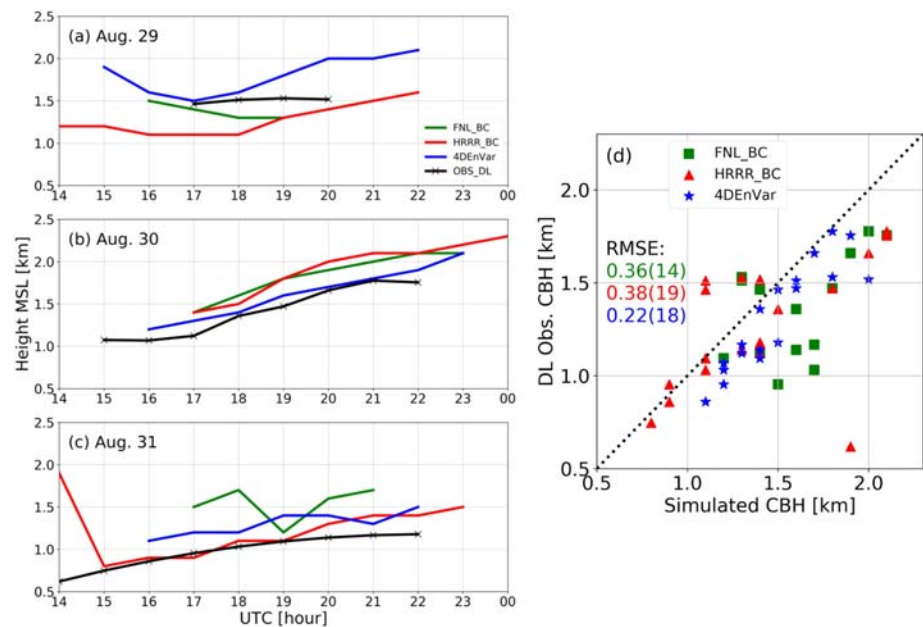
Figure 17. Comparison of domain 4 simulated area-mean CBH (km) from experiments FNL_BC, HRRR_BC, and 4DEnVar and corresponding observation by network of Doppler lidars near ARM SGP site. The time-height plots for the cases of 29, 30, and 31 August are illustrated in (a), (b), and (c), respectively. Scatter plot that includes all valid data points from 3-day period is given in (d) with colors and marker styles for visual identification. The RMSE (km) and number of sampling hours for each experiment are also denoted by colored texts in the plot.

similar temporal variation in CF and LCF with peak values around local noon (18 UTC) that is also similar to observed, the CF from HRRR_BC gradually increases with time. For 31 August, all three simulations reproduce observed CF trend that increases with time (Figure 16c) and HRRR_BC is the closest to the observations. 4DEnVar simulates a somewhat wider extent of clouds than FNL_BC after 18 UTC in the afternoon which is closer to the observations. Figure 16d shows that HRRR_BC overestimates liquid clouds before noon, whereas both FNL_BC and 4DEnVar simulate similar cloud cover to the observations. In the afternoon, three simulations tend to have a much slower rate of decay than observed.

Another metric for model assessment is the CBH as it's also a quasi-observation of the daytime variation in the CBL height. The network of DLs (locations are denoted in Figure 1) can detect ShCu CBH fairly accurately, provides spatial variability, and has much higher temporal resolution than the available radiosondes. Therefore, the DL-estimated CBH is used as the reference for model assessment. Simulated and DL-observed CBH variations for each day are depicted in Figure 17. The simulated CBH is computed by taking the height of lowest model level where area-mean (see the area depicted by black square in Figures 8a and 8b) cloud water mixing ratio exceeds $0.002 \, \text{g kg}^{-1}$. The observed CBH is obtained by least square fitting of all measured data points from five DLs in each day.

The observed diurnal variation of ShCu CBHs (black lines in Figures 17a–17c) shows that there is clear day-to-day variability, corresponding to day-to-day variability in boundary layer evolution. While the lines of observed ShCu CBH on both 30 and 31 August (Figures 17b and 17c) exhibit a rising trend during the day, a much shorter fitted curve of CBH is obtained for 29 August as the measured CBHs among five lidars are unavailable for longer periods. To further assess these simulated results, a scatter plot is shown in Figure 17d that includes all data points of each experiment when both simulated and observed values are valid at a particular time. The root mean square error (RMSE) for each experiment is also listed in Figure 17d. 4DEnVar has the lowest overall error in simulated CBH (0.22 km) among the three experiments with 18 valid samples, suggesting that current DA configuration does have a positive impact on simulation of ShCu populations. On the other hand, FNL_BC has an error of 0.36 km computed from 14 data points, and HRRR_BC on average has an error of 0.38 km from 19 samples.

5.3. Assimilation of Hydrometeor Variables

As shown in section 4.1, the assimilation of hydrometeors in the HRRR model that provides the initial and boundary conditions for the HRRR_BC experiment does constrain the presence of clouds in the model for the first few simulation hours (Figure 4). This result has motivated ongoing research to assimilate gridded retrieval of CWP from GOES-13 measurements along with other observations to further improve the prediction of clouds. Similar to the approach used by Jones et al. (2016) and Chen et al. (2015), the retrieved cloud base and top heights will be supplemented to define the altitude and depth of cloud fields. However, recent studies (Auligné et al., 2010; Wu et al., 2016) describe imbalances between ingested hydrometeors and other prognostic variables that may induce abrupt model adjustments which subsequently negatively impact cloud predictions. Hence, it is necessary to use an effective methodology that updates hydrometeors with smaller imbalances among multiple variables.

5.4. PBL Parameterization Update in Latest HRRR

We previously showed the difficulty of the operational version 2 of HRRR in generating the observed ShCu populations (Figure 6), despite the fact that our HRRR_BC simulation had some success in reproducing the observed cloud fields. This may be due in part to the boundary layer and shallow convective parameterizations within that version of the HRRR. The PBL scheme of the operational HRRR system has been updated to the EDMF version of the MYNN parameterization (MYNN-EDMF; Angevine et al., 2018). In MYNN-EDMF, non-local CBL vertical transport is modeled by an ensemble of subgrid plumes, which enables dry convective turbulence and ShCu to be represented in a unified and physically realistic manner. It is anticipated that such a scheme in an operational system should lead to improved predictions of ShCu populations in scenarios such as those observed on 30 August 2016.

5.5. Potential Application of Kilometer-Scale DA in LES Modeling

Limited by computational resources, conventional LES simulation is usually conducted with a relatively small domain (<100 km wide). Hence, it is common in LES modeling to apply a periodic assumption for the lateral boundaries and prescribe large-scale forcing as domain-averaged profiles. However, as computational power becomes more affordable, it is also possible to use initial and boundary conditions from kilometer-scale models constrained by DA to drive the LES as demonstrated in Haupt et al. (2019) in which they concluded that the major challenge in the microscale simulation for the need of wind industry is to capture the timing of dynamic events in mesoscale.

The framework of kilometer-scale DA demonstrated in this study shows the potential of multiscale DA that enables generation of fine-scale (~1-km resolution) analyses at arbitrary frequencies. Based on the evaluation described in the previous sections, the kilometer-scale simulation is well constrained by observations and could be used to drive LES modeling, such as the simulation described in Fast et al. (2019). To do so, an option is to nest LES domain(s) in the kilometer-scale DA domain so that the DA domain could provide more realistic boundary conditions for free-running inner LES domain(s). Similar work can be also done by taking advantage of the "ndown" function implemented in WRF model. This process acquires inputs from the DA simulation at an arbitrary frequency as long as the domain extent of DA simulation is larger than the LES domain. The initial conditions for WRF LES domain are obtained by interpolation of DA simulation domain at a given time, whereas the boundary conditions are linearly interpolated in time. Hence, the more frequent inputs from DA analyses it has, the more accurate boundary condition will be generated.

6. Summary

In this study, we use an observation-constrained cloud-system resolving model to simulate continental ShCu cloud populations observed on 30 August 2016 during the HI-SCALE field campaign. WRF model forecasts are optimized by assimilating observations including NCEP operational data sets and boundary layer measurements collected near ARM SGP site over north-central Oklahoma with a GSI-based 4DEnVar hybrid technique. To understand the impact of DA on prediction of shallow convective clouds evolution, three additional experiments are conducted, including (1) FNL_BC, which initializes WRF model with FNL as initial and boundary conditions but without DA; (2) HRRR_BC, which initializes WRF by HRRR analyses and without any DA; and (3) 4DEnVar_GDAS, which is similar to 4DEnVar yet only NCEP GDAS data sets are assimilated. The results show that our DA experiment (4DEnVar) reproduces more reasonable

amount of both total and LCFs through daytime over a statewide domain (covering Oklahoma and southern Kansas) than the other experiments against GOES-13 5-km gridded CWP and LWP retrievals, implying that it better captures mesoscale weather systems. Snapshots of simulated cloud fields at 18 UTC from the 4- and 1.33-km domains over north-central Oklahoma reveal that model resolution modulates the number and size of shallow convective cloud population even under very similar ambient conditions. Combining these data with in situ measurements will provide an unprecedented amount data for model evaluation and improved parameterizations.

The mesoscale simulations in this study are evaluated using ARM SGP measurements including DL, radiosonde, RL, and G-1 aircraft. Overall, it shows that the life cycle of ShCu populations is more accurately reproduced by 4DEnVar experiment as it generates shallow convective clouds that are most comparable with time series CBH estimated by DL network. From other evaluations that utilize radiosonde, RL, and G-1 measurements, the 4DEnVar experiment further illustrates how additional measurements made at the ARM SGP site can be used to further constrain models when simulating the evolution of the CBL and shallow clouds. For example, the bias of moisture variations within CBL near the ARM SGP site is smaller in the 4DEnVar experiment than in 4DEnVar_GDAS, which leads to better prediction of the life cycle of shallow convective clouds (Figures 9c and 9d). Of course, the impact of assimilating observational data sets collected near the ARM SGP site is expected to be more significant within a confined region near the site. However, assimilation of surface measurements of Oklahoma Mesonet that has a wider data coverage further extends model constraint in space as it shows clear improvement in predicting precipitation over a larger domain (Figures 6d and 6e), which is also mentioned in Schenkman et al. (2011). Our results suggest that additional measurements of lower-tropospheric moisture profiles at a higher spatiotemporal resolution than the radiosonde network could improve boundary layer and forecasts of shallow convections and subsequently deeper precipitating convective cloud systems over the central United States (Coniglio et al., 2019; National Research Council, 2009).

Some of the obstacles and opportunities of representing shallow convective clouds and their transition to deeper, precipitating convection by cloud-system resolving models are discussed. The likelihood of routine microscale simulations in the near future suggests that additional research on the best approach in coupling kilometer-scale DA to LES modeling is needed. Uncertainties in microscale predictions could be reduced by taking advantage of an increasing number of remote sensing and in situ observations and advanced assimilation techniques in larger-scale mesoscale models.

Data Availability Statement

The GOES data are downloaded from ARM website (https://www.arm.gov/capabilities/vaps/visst). The WRF community model is available from the National Center for Atmospheric Research (NCAR; http://www2.mmm.ucar.edu/wrf/users/). The HRRR analysis and forecast products were obtained from the HRRR archive at University of Utah (http://hrrr.chpc.utah.edu/; doi: 10.7278/S5JQ0Z5B). HI-SCALE data used in this manuscript are freely available from the ARM data archive (https://www.arm.gov/data and https://www.arm.gov/research/campaigns/aaf2016hiscale). The WRF model outputs generated by the simulations in this study are saved on PNNL's long-term storage system, called Aurora (rc-support@pnnl.gov).

References

Adam, S., Behrendt, A., Schwitalla, T., Hammann, E., & Wulfmeyer, V. (2016). First assimilation of temperature lidar data into an NWP model: Impact on the simulation of the temperature field, inversion strength and PBL depth. *Quarterly Journal of the Royal Meteorological Society*, *142*(700), 2882–2896. https://doi.org/10.1002/qj.2875

Alapaty, K., Seaman, N. L., Niyogi, D. S., & Hanna, A. F. (2001). Assimilating surface data to improve the accuracy of atmospheric boundary layer simulations. *Journal of Applied Meteorology*, *40*(11), 2068–2082. https://doi.org/10.1175/1520-0450(2001)040<2068:ASDTIT>2.0.CO;2

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. https://doi.org/10.1175/2009bams2618.1

Angevine, W. M., Olson, J., Kenyon, J., Gustafson, W. I. Jr., Endo, S., Suselj, K., & Turner, D. D. (2018). Shallow cumulus in WRF parameterizations evaluated against LASSO large-eddy simulations. *Monthly Weather Review*, *146*(12), 4303–4322. https://doi.org/10.1175/MWR-D-18-0115.1

Auligné, T., Lorenc, A., Michel, Y., Montmerle, T., Jones, A., Hu, M., & Dudhia, J. (2010). Toward a new cloud analysis and prediction system. *Bulletin of the American Meteorological Society*, *92*(2), 207–210. https://doi.org/10.1175/2010BAMS2978.1

Bannister, R. N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 607–633. https://doi.org/10.1002/qj.2982

Barker, D., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., et al. (2012). The weather research and forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bulletin of the American Meteorological Society*, *93*(6), 831–843. https://doi.org/10.1175/bams-d-11-00167.1

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The rapid refresh. *Monthly Weather Review*, *144*(4), 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Berg, E. I., Gustafso, W. I., Kassianov, E. I., & Deng, L. (2013). Evaluation of a modified scheme for shallow convection: Implementation of CuP and case studies. *Monthly Weather Review*, *141*(1), 134–147. https://doi.org/10.1175/MWR-D-12-00136.1

Berg, E. I., Kassianov, L. K., Long, C. N., & Mills, D. L. (2011). Surface summertime radiative forcing by shallow cumuli at the atmospheric radiation measurement Southern Great Plains site. *Journal of Geophysical Research*, *116*, D01202. https://doi.org/10.1029/2010JD04593

Berg, L. K., & Kassianov, E. I. (2008). Temporal variability of fair-weather cumulus statistics at the ACRF SGP site. *Journal of Climate*, *21*(13), 3344–3358. https://doi.org/10.1175/2007JCLI2266.1

Chang, S.-F., Liou, Y.-C., Sun, J., & Tai, S.-L. (2015). The implementation of the ice-phase microphysical process into a four-dimensional Variational Doppler Radar Analysis System (VDRAS) and its impact on parameter retrieval and quantitative precipitation nowcasting. *Journal of the Atmospheric Sciences*, *73*(3), 1015–1038. https://doi.org/10.1175/JAS-D-15-0184.1

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State—NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*, 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Chen, Y., Wang, H., Min, J., Huang, X.-Y., Minnis, P., Zhang, R., et al. (2015). Variational assimilation of cloud liquid/ice water path and its impact on NWP. *Journal of Applied Meteorology and Climatology*, *54*(8), 1809–1825. https://doi.org/10.1175/JAMC-D-14-0243.1

Clothiaux, E. E., Miller, M. A., Perez, R. C., Turner, D. D., Moran, K. P., Martner, B. E., et al. (2001) The ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP), Pacific Northwest Natl. Lab. https://www.arm.gov/publications/tech_reports/arm-vap-002-1.pdf

Coniglio, G., Romine, S., Turner, D. D., & Torn, R. D. (2019). Impacts of targeted AERI and Doppler lidar wind retrievals on short-term forecasts of the initiation and early evolution of thunderstorms. *Monthly Weather Review*, *147*(4), 1149–1170. https://doi.org/10.1175/MWR-D-18-0351.1

Fast, J. D., Berg, L. K., Alexander, L., Bell, D., D'Ambro, E., Hubbe, J., et al. (2018). Overview of the HI-SCALE field campaign: A new perspective on shallow convective clouds. *Bulletin of the American Meteorological Society*, *100*(5), 821–840. https://doi.org/10.1175/BAMS-D-18-0030.1

Fast, J. D., Berg, L. K., Feng, Z., Mei, F., Newsom, R., Sakaguchi, K., & Xiao, H. (2019). The impact of variable land-atmosphere coupling on convective cloud populations observed during the 2016 HI-SCALE field campaign. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2629–2654. https://doi.org/10.1029/2019ms001727

Fast, J. D., Gustafson, W. I., Chapman, E. G., Easter, R. C., Rishel, J. P., Zaveri, R. A., et al. (2011). The aerosol modeling testbed: A community tool to objectively evaluate aerosol process modules. *Bulletin of the American Meteorological Society*, *92*(3), 343–360. https://doi.org/10.1175/2010BAMS2868.1

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*(14), 1693. https://doi.org/10.1029/2002GL015311

Gustafson, W. I., Vogelmann, A. M., Li, Z., Cheng, X., Dumas, K. K., Endo, S., et al. (2020). The large-eddy simulation (LES) atmospheric radiation measurement (ARM) symbiotic simulation and observation (LASSO) activity for continental shallow convection. *Bulletin of the American Meteorological Society*, *101*(4), E462–E479. https://doi.org/10.1175/bams-d-19-0065.1

Hamill, T. M., Whitaker, J. S., Kleist, D. T., Fiorino, M., & Benjamin, S. G. (2011). Predictions of 2010's tropical cyclones using the GFS and ensemble-based data assimilation methods. *Monthly Weather Review*, *139*(10), 3243–3247. https://doi.org/10.1175/mwr-d-11-00079.1

Haupt, S. E., Kosovic, B., Shaw, W., Berg, L. K., Churchfield, M., Cline, J., et al. (2019). On bridging a modeling scale gap: Mesoscale to microscale coupling for wind energy. *Bulletin of the American Meteorological Society*, *100*(12), 2533–2550. https://doi.org/10.1175/bams-d-18-0033.1

Hong, S.-Y., & Lim, J.-O. J. (2006). The WRF single-moment 6-class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, *42*, 129–151.

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Hu, M., Benjamin, S. G., Ladwig, T. T., Dowell, D. C., Weygandt, S. S., Alexander, C. R., & Whitaker, J. S. (2017). GSI three-dimensional ensemble–variational hybrid data assimilation using a global ensemble for the regional rapid refresh model. *Monthly Weather Review*, *145*(10), 4205–4225. https://doi.org/10.1175/mwr-d-16-0418.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD009944

Janjić, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, *122*(5), 927–945. https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2

Janjić, Z. I. (2001). Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. *Met Off. Note*, *437*, 61. https://repository.library.noaa.gov/view/noaa/11409

Jimenez, P. A., Hacker, J. P., Dudhia, J., Haupt, S. E., Ruiz-Arias, J. A., Gueymard, C. A., et al. (2016). WRF-Solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction. *Bulletin of the American Meteorological Society*, *97*(7), 1249–1264. https://doi.org/10.1175/bams-d-14-00279.1

Johnson, A., Wang, X., Carley, J. R., Wicker, L. J., & Karstens, C. (2015). A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Monthly Weather Review*, *143*(8), 3087–3108. https://doi.org/10.1175/MWR-D-14-00345.1

Jones, T. A., Knopfmeier, K., Wheatley, D., Creager, G., Minnis, P., & Palikonda, R. (2016). Storm-scale data assimilation and ensemble forecasting with the NSSL experimental warn-on-forecast system. Part II: Combined Radar and Satellite Data Experiments. *Weather and Forecasting*, *31*(1), 297–327. https://doi.org/10.1175/WAF-D-15-0107.1

Khairoutdinov, M., & Randall, D. (2006). High-resolution simulation of shallow-to-deep convection transition over land. *Journal of the Atmospheric Sciences*, *63*(12), 3421–3436. https://doi.org/10.1175/JAS3810.1

Kuang, Z., & Bretherton, C. S. (2006). A mass-flux scheme view of a high-resolution simulation of a transition from shallow to deep cumulus convection. *Journal of the Atmospheric Sciences*, *63*(7), 1895–1909. https://doi.org/10.1175/JAS3723.1

Lareau, N. P., Zhang, Y., & Klein, S. A. (2018). Observed boundary layer controls on shallow cumulus at the ARM Southern Great Plains site. *Journal of the Atmospheric Sciences*, *75*(7), 2235–2255. https://doi.org/10.1175/JAS-D-17-0244.1

LeMone, M. A., Tewari, M., Chen, F., & Dudhia, J. (2013). Objectively determined fair-weather CBL depths in the ARW-WRF model and their comparison to CASES-97 observations. *Monthly Weather Review*, *141*(1), 30–54. https://doi.org/10.1175/MWR-D-12-00106.1

Li, Z., Feng, S., Liu, Y., Lin, W., Zhang, M., Toto, T., et al. (2015). Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation: Fine resolution analysis and forcing. *Journal of Geophysical Research: Atmospheres*, *120*, 654–666. https://doi.org/10.1002/2014JD022245

Lin, Y., & Mitchell, K. E. (2005) The NCEP stage II/IV hourly precipitation analyses: Development and applications. Pre-prints,*19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. [available online at https://ams.confex.com/ams/pdfpapers/83847.pdf.]

Liu, S., & Liang, X.-Z. (2010). Observed diurnal cycle climatology of planetary boundary layer height. *Journal of Climate*, *23*(21), 5790–5809. https://doi.org/10.1175/2010jcli3552.1

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., et al. (2007). Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, *24*(3), 301–321. https://doi.org/10.1175/JTECH1976.1

Meng, Z., & Zhang, F. (2008). Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Monthly Weather Review*, *136*(2), 522–540. https://doi.org/10.1175/2007mwr2106.1

Minnis, P., Nguyen, L., Palikonda, R., Heck, P. W., Spangenberg, D. A., Doelling, D. R., et al. (2008). Near-real time cloud retrievals from operational and research meteorological satellites. In R. H. Picard, A. Comeron, K. Schäfer, A. Amodeo, & M. van Weele (Eds.), *SPIE Remote Sensing* (710703). United Kingdom: Cardiff, Wales.

Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., et al. (2011). CERES Edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(11), 4374–4400. https://doi.org/10.1109/TGRS.2011.2144601

Morcrette, C. J., Van Weverberg, K., Ma, H.-Y., Ahlgrimm, M., Bazile, E., Berg, L. K., et al. (2018). Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the southern great plains. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2655–2683. https://doi.org/10.1002/2017JD027199

Morrison, H., Curry, J. A., & Khvorostyanov, V. I. (2005). A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *Journal of the Atmospheric Sciences*, *62*(6), 1665–1677. https://doi.org/10.1175/JAS3446.1

Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan Series II*, *87*(5), 895–912. https://doi.org/10.2151/jmsj.87.895

National Research Council (2009). *Observing weather and climate from the ground up: A nationwide network of networks*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12540

Neggers, R. A. J., Neelin, J. D., & Stevens, B. (2007). Impact mechanisms of shallow cumulus convection on tropical climate dynamics. *Journal of Climate*, *20*(11), 2623–2642. https://doi.org/10.1175/JCLI4079.1

Newsom, R. K. (2009). *Raman Lidar (RL) Handbook; DOE/SC-ARM/TR-038*. Washington, DC, USA: Department of Energy. https://www.arm.gov/publications/tech_reports/handbooks/rl_handbook.pdf

Niu, G.-Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*, D12109. https://doi.org/10.1029/2010JD015139

Olson, J. B., Kenyon, J. S., Angevine, W. A., Brown, J. M., Pagowski, M., & Suselj, K. (2019) A description of the MYNN-EDMF scheme and the coupling to other components in WRF-ARW. NOAA Technical Memorandum OAR GSD-61. https://repository.library.noaa.gov/view/noaa/19837

Ruggiero, F. H., Sashegyi, K. D., Madala, R. V., & Raman, S. (1996). The use of surface observations in four-dimensional data assimilation using a mesoscale model. *Monthly Weather Review*, *124*(5), 1018–1033. https://doi.org/10.1175/1520-0493(1996)124<1018:TUOSOI>2.0.CO;2

Schenkman, A. D., Xue, M., Shapiro, A., Brewster, K., & Gao, J. (2011). Impact of CASA radar and Oklahoma Mesonet data assimilation on the analysis and prediction of tornadic meso-vortices in an MCS. *Monthly Weather Review*, *139*(11), 3422–3445. https://doi.org/10.1175/MWR-D-10-05051.1

Schmid, B., Tomlinson, J. M., Hubbe, J. M., Comstock, J. M., Mei, F., Chand, D., et al. (2013). The DOE ARM Aerial Facility. *Bulletin of the American Meteorological Society*, *95*(5), 723–742. https://doi.org/10.1175/BAMS-D-13-00040.1

Schwartz, C. S., Liu, Z., & Huang, X.-Y. (2015). Sensitivity of limited-area hybrid variational-ensemble analyses and forecasts to ensemble perturbation resolution. *Monthly Weather Review*, *143*(9), 3454–3477. https://doi.org/10.1175/mwr-d-14-00259.1

Shao, H., Derber, J., Huang, X. Y., Hu, M., Newman, K., Stark, D., et al. (2015). Bridging research to operations transitions: Status and plans of community GSI. *Bulletin of the American Meteorological Society*, *97*(8), 1427–1440. https://doi.org/10.1175/BAMS-D-13-00245.1

Sisterson, D. L., Peppler, R. A., Cress, T. S., Lamb, P. J., & Turner, D. D. (2016). The ARM Southern Great Plains (SGP) site. *Meteorological Monographs*, *57*, 6.1–6.14. https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1

Skamarock, C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., & Powers, J. G. (2008) A description of the advanced research WRF version 3. https://doi.org/10.5065/D68S4MVH

Stauffer, D. R., Seaman, N. L., & Binkowski, F. S. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model part II: Effects of data assimilation within the planetary boundary layer. *Monthly Weather Review*, *119*, 734–754. https://doi.org/10.1175/1520-0493(1991)119<0734:UOFDDA>2.0.CO;2

Sun, J., & Wang, H. (2012). Radar data assimilation with WRF 4D-Var. Part II: Comparison with 3D-Var for a squall line over the U.S. Great Plains. *Monthly Weather Review*, *141*, 2245–2264. https://doi.org/10.1175/MWR-D-12-00169.1

Tai, S.-L., Liou, Y.-C., Sun, J., Chang, S.-F., & Kuo, M.-C. (2011). Precipitation forecasting using Doppler radar data, a cloud model with adjoint, and the weather research and forecasting model: Real case studies during SoWMEX in Taiwan. *Weather and Forecasting*, *26*(6), 975–992. https://doi.org/10.1175/WAF-D-11-00019.1

Tai, Y., Liou, Y. C., Sun, J., & Chang, S. F. (2017). The development of a terrain-resolving scheme for the forward model and its adjoint in the four-dimensional Variational Doppler Radar Analysis System (VDRAS). *Monthly Weather Review*, *145*(1), 289–306. https://doi.org/10.1175/MWR-D-16-0092.1

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*(8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

Wagner, T. J., Turner, D. D., Berg, L. K., & Krueger, S. K. (2013). Ground-based remote retrievals of cumulus entrainment rates. *Journal of Atmospheric and Oceanic Technology*, *30*(7), 1460–1471. https://doi.org/10.1175/JTECH-D-12-00187.1

Wang, X., & Lei, T. (2014). GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Monthly Weather Review*, *142*(9), 3303–3325. https://doi.org/10.1175/MWR-D-13-00303.1

Wang, X., Parrish, D., Kleist, D., & Whitaker, J. (2013). GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Monthly Weather Review*, *141*(11), 4098–4117. https://doi.org/10.1175/mwr-d-12-00141.1

Wu, C.-M., Stevens, B., & Arakawa, A. (2009). What controls the transition from shallow to deep convection? *Journal of the Atmospheric Sciences*, *66*(6), 1793–1806. https://doi.org/10.1175/2008JAS2945.1

Wu, T.-C., Zupanski, M., Grasso, L. D., Brown, P. J., Kummerow, C. D., & Knaff, J. A. (2016). The GSI capability to assimilate TRMM and GPM hydrometeor retrievals in HWRF. *Quarterly Journal of the Royal Meteorological Society*, *142*(700), 2768–2787. https://doi.org/10.1002/qj.2867

Wulfmeyer, V., Bauer, H.-S., Grzeschik, M., Behrendt, A., Vandenberghe, F., Browell, E. V., et al. (2006). Four-dimensional variational assimilation of water vapor differential absorption lidar data: The first case study within IHOP_2002. *Monthly Weather Review*, *134*(1), 209–230. https://doi.org/10.1175/MWR3070.1

Xiao, Q., & Sun, J. (2007). Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP_2002. *Monthly Weather Review*, *135*(10), 3381–3404. https://doi.org/10.1175/MWR3471.1

Zhang, Y., & Klein, S. A. (2010). Mechanisms affecting the transition from shallow to deep convection over land: Inferences from observations of the diurnal cycle collected at the ARM Southern Great Plains site. *Journal of the Atmospheric Sciences*, *67*(9), 2943–2959. https://doi.org/10.1175/2010JAS3366.1

Zhang, Y. S., & Klein, S. A. (2012). Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the ARM Southern Great Plains site. *Journal of the Atmospheric Sciences*, *70*(4), 1297–1315. https://doi.org/10.1175/JAS-D-12-0131.1

Zhu, K., Pan, Y., Xue, M., Wang, X., Whitaker, J. S., Benjamin, S. G., et al. (2013). A regional GSI-based ensemble Kalman filter data assimilation system for the rapid refresh configuration: Testing at reduced resolution. *Monthly Weather Review*, *141*(11), 4118–4139. https://doi.org/10.1175/MWR-D-13-00039.1